

STIC-Biotech/ChemLib

31155

From: Brunovskis, Peter
Sent: Tuesday, December 12, 2000 6:45 PM
To: STIC-Biotech/ChemLib
Subject: sequence search on 09/051,034

Please search SEQ ID NOs:11, -12, and -13 of 09/051,034 against the commercial and interference databases using standard search parameters.

Please print results on laser printer and send to Peter Brunovskis, CM1-12E05. GAU 1632. 305-2471.

Thanks

Point of Contact:
Mona Smith
Technical Info. Specialist
CM1 12C14 Tel: 308-3278

Point of Contact:
Mons Smith
Technical Info. Specialist
CMT 1SC14 Tel: 308-3278

THIS PAGE BLANK (USPTO)

Scientific and Technical Information Center

31155

Requester's Full Name: _____ Examiner #: _____ Date: _____
 Art Unit: _____ Phone Number 30 _____ Serial Number: _____
 Mail Box and Bldg/Room Location: _____ Results Format Preferred (circle): PAPER DISK E-MAIL

If more than one search is submitted, please prioritize searches in order of need.

 Please provide a detailed statement of the search topic, and describe as specifically as possible the subject matter to be searched. Include the elected species or structures, keywords, synonyms, acronyms, and registry numbers, and combine with the concept or utility of the invention. Define any terms that may have a special meaning. Give examples or relevant citations, authors, etc, if known. Please attach a copy of the cover sheet, pertinent claims, and abstract.

Title of Invention: _____

Inventors (please provide full names): _____

Earliest Priority Filing Date: _____

**For Sequence Searches Only* Please include all pertinent information (parent, child, divisional, or issued patent numbers) along with the appropriate serial number.*

STAFF USE ONLY

Searcher: M. Smith
 Searcher Phone #: _____
 Searcher Location: _____
 Date Searcher Picked Up: 12-14-00
 Date Completed: 12-15-00
 Searcher Prep & Review Time: 15
 Clerical Prep Time: _____
 Online Time: 10

Type of Search

NA Sequence (#) _____
 AA Sequence (#) 3
 Structure (#) _____
 Bibliographic _____
 Litigation _____
 Fulltext _____
 Patent Family _____
 Other _____

Vendors and cost where applicable

STN _____
 Dialog _____
 Questel/Orbit _____
 Dr.Link _____
 Lexis/Nexis _____
 Sequence Systems _____
 WWW/Internet _____
 Other (specify) _____

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 14, 2000, 10:10:26 ; Search time 36.07 Seconds
(without alignments)
5.688 Million cell updates/sec

Title: US-09-051-034A-11

Perfect score: 31

Sequence: 1 MNVGR 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A.GeneSeq_36:*

- 1: /SIDSI/gcgdata/geneSeq/geneSeq/AA1980.DAT:*
- 2: /SIDSI/gcgdata/geneSeq/geneSeq/AA1981.DAT:*
- 3: /SIDSI/gcgdata/geneSeq/geneSeq/AA1982.DAT:*
- 4: /SIDSI/gcgdata/geneSeq/geneSeq/AA1983.DAT:*
- 5: /SIDSI/gcgdata/geneSeq/geneSeq/AA1984.DAT:*
- 6: /SIDSI/gcgdata/geneSeq/geneSeq/AA1985.DAT:*
- 7: /SIDSI/gcgdata/geneSeq/geneSeq/AA1986.DAT:*
- 8: /SIDSI/gcgdata/geneSeq/geneSeq/AA1987.DAT:*
- 9: /SIDSI/gcgdata/geneSeq/geneSeq/AA1988.DAT:*
- 10: /SIDSI/gcgdata/geneSeq/geneSeq/AA1989.DAT:*
- 11: /SIDSI/gcgdata/geneSeq/geneSeq/AA1990.DAT:*
- 12: /SIDSI/gcgdata/geneSeq/geneSeq/AA1991.DAT:*
- 13: /SIDSI/gcgdata/geneSeq/geneSeq/AA1992.DAT:*
- 14: /SIDSI/gcgdata/geneSeq/geneSeq/AA1993.DAT:*
- 15: /SIDSI/gcgdata/geneSeq/geneSeq/AA1994.DAT:*
- 16: /SIDSI/gcgdata/geneSeq/geneSeq/AA1995.DAT:*
- 17: /SIDSI/gcgdata/geneSeq/geneSeq/AA1996.DAT:*
- 18: /SIDSI/gcgdata/geneSeq/geneSeq/AA1997.DAT:*
- 19: /SIDSI/gcgdata/geneSeq/geneSeq/AA1998.DAT:*
- 20: /SIDSI/gcgdata/geneSeq/geneSeq/AA1999.DAT:*
- 21: /SIDSI/gcgdata/geneSeq/geneSeq/AA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	100.0	342	19	Porcine alpha-1,3-galactosyl transferase
2	31	100.0	354	15	Porcine alpha-1,3-galactosyl transferase
3	31	100.0	359	15	Porcine alpha-1,3-galactosyl transferase
4	31	100.0	359	17	Porcine alpha-1,3-galactosyl transferase
5	31	100.0	363	19	Porcine alpha-1,3-galactosyl transferase
6	31	100.0	371	16	Porcine alpha-1,3-galactosyl transferase
7	31	100.0	371	16	Porcine alpha-1,3-galactosyl transferase
8	31	100.0	375	19	Porcine alpha-1,3-galactosyl transferase
9	28	90.3	100	20	Secreted protein n
10	28	90.3	376	16	Marmoset alpha-1,3
11	28	90.3	394	12	GDP-Fuc: [beta-D-Ga
12	28	90.3	394	15	A glycosyltransferase

13	28	90.3	394	18	W13639
14	27	87.1	94	18	W27869
15	27	87.1	165	19	W38713
16	26	83.9	332	18	W27808
17	26	83.9	118	19	W59167
18	26	83.9	123	14	R43827
19	26	83.9	142	18	W09816
20	26	83.9	274	18	W09819
21	26	83.9	282	18	W35564
22	26	83.9	282	18	W09818
23	26	83.9	289	18	W35563
24	26	83.9	289	21	V67583
25	26	83.9	385	21	V78855
26	26	83.9	532	18	W35565
27	26	83.9	972	20	V6212
28	26	83.9	972	21	V43099
29	25	80.6	146	15	R58554
30	25	80.6	161	19	W98472
31	25	80.6	171	15	R58553
32	25	80.6	175	15	R58556
33	25	80.6	198	16	R80781
34	25	80.6	198	16	R70826
35	25	80.6	198	19	W75716
36	25	80.6	198	19	W53032
37	25	80.6	198	21	V90415
38	25	80.6	198	21	V32338
39	25	80.6	208	13	R24357
40	25	80.6	208	15	R58555
41	25	80.6	218	18	W20717
42	25	80.6	227	9	P82079
43	25	80.6	227	11	R05231
44	25	80.6	228	18	W23620
45	25	80.6	228	18	W23626

ALIGNMENTS

RESULT 1

W49689 standard; Protein; 342 AA.

W49689:

10-NOV-1998 (first entry)

Porcine alpha-1,3-galactosyl transferase isoform 4.

Isoform: porcine; enzyme: alpha-1,3-galactosyl transferase; galactose: sugar; N-acetylglucosamine; glycoprotein; glycolipid; antibody: pig; graft tissue rejection; organ transplantation; xenotransplant.

Sus scrofa.

FR2751346-A1.

23-JAN-1998.

19-JUL-1996: 96FR-0009077.

19-JUL-1996: 96FR-0009077.

19-JUL-1996: 96FR-0009077.

(INRM) INSERM INST NAT SANTE & RECH MEDICALE.

Pourcel C, Soullieu JP, Vanhove B:

WPI: 1998-112876/11.

N-PSDB: V49456.

Transgenic non-human donors of organs for human recipients - containing DNA encoding antibodies that inhibit graft rejection

Claim 4; Page 43-44; 71pp; French.

XX This sequence represents isoform 4 of the porcine enzyme
 CC alpha-1,3-galactosyl transferase (alpha-1,3-GT). The enzyme catalyses
 CC the attachment of a galactose sugar molecule on the N-acetylglucosamine
 CC moiety found on surface glycoproteins and glycolipids. These sugar
 CC molecules are partly responsible for raising anti-graft antibodies, which
 CC lead to graft tissue rejection. The invention relates to a method of
 CC inhibiting the graft rejection mechanism by introducing the sequence
 CC encoding an antibody targeted to alpha-1,3-GT into the cells of animal,
 CC especially a pig, from whom organs may be used for xenotransplants.
 CC Neutralisation of the alpha-1,3-GT leads to tissues or organs lacking
 CC the galactose on the glycoproteins and glycolipids, thus preventing
 CC induction of the rejection response.
 CC
 CC Sequence 342 AA;

Query Match 100.0%; Score 31; DB 19; Length 342;
 Best Local Similarity 100.0%; Pred. No. 8.3;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVGR 6
 |||||
 Db 5 mnvgr 10

RESULT 2

ID W09688 standard; Protein; 354 AA.

AC W09688;

DT 10-NOV-1998 (first entry)

DE Porcine alpha-1,3-galactosyl transferase isoform 3.

XX Isoform; porcine; enzyme; alpha-1,3-galactosyl transferase; galactose;
 KW Sugar; N-acetylglucosamine; glycoprotein; glycolipid; antibody; pig;
 KW graft tissue rejection; organ transplantation; xenotransplant.

XX Sus scrofa.

XX P02751346-A1.

PD 23-JAN-1998.

PF 10-JUL-1996; 96FR-0009077.

PR 19-JUL-1996; 96FR-0009077.

PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

PI Pourcel C, Soullion JP, Vanhove B;

DR WPI; 1998-112876/11.

DR NPSDB; V49455.

PT Transgenic non-human donors of organs for human recipients -
 CC containing DNA encoding antibodies that inhibit graft rejection

PS Claim 4; Page 39-41; 71pp: French.

XX This sequence represents isoform 3 of the porcine enzyme
 CC alpha-1,3-galactosyl transferase (alpha-1,3-GT). The enzyme catalyses
 CC the attachment of a galactose sugar molecule on the N-acetylglucosamine
 CC moiety found on surface glycoproteins and glycolipids. These sugar
 CC molecules are partly responsible for raising anti-graft antibodies, which
 CC lead to graft tissue rejection. The invention relates to a method of
 CC inhibiting the graft rejection mechanism by introducing the sequence
 CC encoding an antibody targeted to alpha-1,3-GT into the cells of animal,
 CC especially a pig, from whom organs may be used for xenotransplants.
 CC Neutralisation of the alpha-1,3-GT leads to tissues or organs lacking
 CC the galactose on the glycoproteins and glycolipids, thus preventing

CC Induction of the rejection response.

XX Sequence 354 AA;

Query Match 100.0%; Score 31; DB 19; Length 354;
 Best Local Similarity 100.0%; Pred. No. 8.7;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVGR 6
 |||||
 Db 5 mnvgr 10

RESULT 3

ID R62508 standard; Protein; 359 AA.

AC R62508;

DT 26-JUN-1995 (first entry)

DE Galactosyl transferase clone product.

XX Gal-alpha (1,3) galactosyl transferase; xenograft; transplant;
 KW rejection.

OS Sus scrofa domestica.

PN W09421799-A.

PD 29-SEP-1994.

PF 15-MAR-1994; 94WO-AU00126.

PR 16-MAR-1993; 93AU-0007854.

PA (AUST-) AUSTIN RES INST.

PI McKenzie IFC, Sandrin MS;

DR WPI; 1994-317019/39.

DR NPSDB; Q74712.

PT DNA sequences encoding Gal-alpha (1,3)galactosyl transferase -
 CC and clones contg. such sequences are used in xenograft therapies

PS Disclosure; Page 35; 50pp: English.

XX The sequence is that of the product of the porcine Gal-alpha (1,3)
 CC galactosyl transferase gene which produces a Gal epitope on the
 CC surface of porcine cells. This epitope is recognised by antibodies
 CC which are responsible for hyperacute rejection of xenotransplanted
 CC pig cells, tissues and organs.
 CC See also R62307.

XX Sequence 359 AA;

Query Match 100.0%; Score 31; DB 15; Length 359;
 Best Local Similarity 100.0%; Pred. No. 8.8;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVGR 6
 |||||
 Db 1 mnvgr 6

RESULT 4

ID R90573 standard; Protein; 359 AA.

AC R90573;

Query Match 100.0%; Score 31; DB 19; Length 375;
 Best Local Similarity 100.0%; Pred. No. 9.2;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVGR 6
 |||||
 Db 5 mvkgr 10

RESULT 9
 Y26039 Y26039 standard; Protein: 100 AA.
 XX
 AC Y26039;
 XX
 DT 22-OCT-1999 (first entry)
 XX
 DE Secreted protein nt746_4.
 XX
 KM Secreted protein; cytokine; cell proliferation; immune stimulation;
 KM vaccine; immune suppression; haematopoiesis; tissue growth; activin;
 KM inhibin; chemotaxis; chemokinesis; haemostasis; thrombolytic;
 KM receptor; ligand; anti-inflammatory; cadherin; tumour; gene therapy.
 XX
 OS Homo sapiens.
 XX
 PN WO9928335-A1.
 XX
 PD 10-JUN-1999.
 XX
 PF 02-DEC-1998; 98WO-US25512.
 XX
 PR 30-NOV-1998; 98US-0203106.
 PR 04-DEC-1997; 97US-0067454.
 XX
 PA (GENEY) GENETICS INST INC.
 XX
 PI Agostino MJ, Clark HF, Collins-Racie LA, Evans C;
 PI Reichel K, Jacobs K, Lavallee ER, McCoy JM, Weiberg D;
 PI Steininger RJ, Treacy M, Wong GG;
 XX
 DR WPI: 1999-385352/32.
 XX
 N-PSDB: X80674.
 XX
 PS New polynucleotides encoding secreted human proteins
 C1a1n 34; Page 115; 124pp; English.
 XX
 CC The present sequence is a secreted protein nt746_4, encoded by known
 CC clone nt746_4 (deposited as ATCC 98600) isolated from human adult brain
 CC cDNA library. Recombinant secreted proteins can be produced by
 CC transforming host cells and culturing them under suitable conditions.
 CC The polynucleotide and protein are predicted to have biological
 CC activities which would make them suitable for treating, preventing or
 CC ameliorating medical conditions in humans and animals. Some predicted
 CC biological activities include cytokine and cell proliferation/
 CC differentiation activity, immune stimulating (e.g. as vaccines) or
 CC suppressing activity, haematopoiesis regulating activity, tissue growth
 CC activity, activin/inhibin activity, chemotactic/chemokinetic activity,
 CC hemostatic and thrombolytic activity, receptor/ligand activity, anti-
 CC inflammatory activity, cadherin/tumour invasion suppressor activity, and
 CC tumour inhibition activity. The polynucleotide encoding secreted
 CC protein can be used for gene therapy.
 XX
 SQ Sequence 100 AA;

Query Match 90.3%; Score 28; DB 20; Length 100;
 Best Local Similarity 83.3%; Pred. No. 11;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVGR 6
 |||||
 Db 1 mvkgr 6

RESULT 10
 R80016 R80016 standard; Protein: 376 AA.
 XX
 AC R80016;
 XX
 DT 01-MAY-1996 (first entry)
 XX
 DE Marmoset alpha-1,3-galactosyltransferase.
 XX
 KM Marmoset; alpha-1,3-galactosyltransferase; immune response; glycoprotein;
 KM alpha-galactosyl epitope; cell membrane; virus; phagocytosis; tumour;
 KM antigen processing; leukemia; lymphoma; myeloma; melanoma; carcinoma;
 KM sarcoma; vaccine; opsonisation; glycoprotein; antibody; anti-Gal.
 XX
 OS Callithrix jacchus.
 XX
 PN WO954924-A1.
 XX
 PD 21-SEP-1995.
 XX
 PF 13-MAR-1995; 95WO-US03156.
 XX
 PR 15-MAR-1994; 94US-0213200.
 XX
 PA (UYHA-) UNIV HAHNEMANN & MEDICAL COLLEGE PENNSYL.
 XX
 PI Gallili U, Replik PM;
 XX
 DR WPI: 1995-336816/43.
 DR N-PSDB: T04522.
 XX
 PT Association of an alpha-galactosyl epitope with a tumour or viral
 PT antigen - is administered to anti-gal synthesising animals to induce
 PT an immune response
 XX
 PS Disclosure: Fig 9; 85pp; English.
 XX
 CC The amino acid sequence of the marmoset alpha-1,3-galactosyltransferase.
 CC The enzyme can be used in methods of enhancing an immune response by
 CC associating the alpha-galactosyl epitope with a cell membrane or viral
 CC glycoprotein. The alpha-galactosyl epitope enhances phagocytosis and
 CC subsequent processing of the antigen. The method is useful in the
 CC treatment of tumours e.g. leukemia, lymphoma, myeloma, melanoma,
 CC carcinoma and sarcoma, or for the generation of viral vaccines by
 CC opsonising a viral glycoprotein. The alpha-galactosyl epitope enhances
 CC recognition of the antigen in an animal that synthesises the naturally
 CC occurring antibody - anti-Gal.
 XX
 SQ Sequence 376 AA;

Query Match 90.3%; Score 28; DB 16; Length 376;
 Best Local Similarity 83.3%; Pred. No. 48;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVGR 6
 |||||
 Db 1 mvkgr 6

RESULT 11
 R13750 R13750 standard; Protein: 394 AA.
 XX
 AC R13750;
 XX
 DT 07-NOV-1991 (first entry)

XX GDP-Fuc:[beta-D-Gal(1,4/1,3)]-D-GlcNAc/(Glc)alpha(1,3/1,4)
 DE -fucosyltransferase.
 XX Glycosyltransferase.
 XX Mus musculus.
 OS Mus musculus.
 XX W09112340-A.
 PN W09112340-A.
 PD 22-AUG-1991.
 XX 14-FEB-1991; 91WO-US00899.
 PF 14-FEB-1991; 90US-0627621.
 PR 14-FEB-1990; 90US-0479858.
 PR 14-FEB-1990; 90US-0480133.
 XX (UNMI) UNIV OF MICHIGAN.
 PA (UNMI) UNIV OF MICHIGAN.
 XX Lowe JB;
 PI WPI: 1991-267151/36.
 DR N-PSDB: Q13331.
 XX Isolation of gene conveying post-translational characteristic -
 PT e.g. the presence of soluble or membrane bound oligo or
 PT polysaccharide or glycosyltransferase.
 XX Disclosure; Fig 2; 155pp; English.
 PS The amino acid sequence codes for a protein capable of functioning
 CC as a UDP-Gal:[beta-D-Gal(1,4)]-D-GlcNAc alpha (1,3)galacto-
 CC syltransferase.. The products of this enzyme, sub-terminal alpha
 CC (1,3) and alpha(1,4) fucose residues are used in the post-
 CC translational modification of the oligosaccharides on cell-surface,
 CC intracellular or secreted proteins or lipids. These can be used for
 CC the prodn. of diagnostics and therapeutics. There is a single
 CC transmembrane domain consisting of a 19 amino acid hydrophobic
 CC segment flanked by basic residues and a large (presumably
 CC catalytic) C-terminal domain that would ultimately be targeted to
 CC the lumen of the Golgi. It has two potential N-glycosylation sites
 CC indicating that as with other glycosyltransferases, it may be
 CC synthesised as a glycoprotein. It is representative of a Type II
 CC transmembrane protein. See also R13749-R13752.
 CC
 CC Sequence 394 AA;
 SQ

Query Match 90.3%; Score 28; DB 12; Length 394;
 Best Local Similarity 83.3%; Pred. No. 51;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNVGR 6
 |||||
 Db 36 mnvrgk 41

RESULT 12
 ID R45935 standard; Protein; 394 AA.
 AC R45935;
 XX 26-JUL-1994 (first entry)
 DT A glycosyltransferase.
 DE Glycosyltransferase: fucosyltransferase; GDP-Fuc; in vitro; cell;
 KM surface; oligosaccharide.
 XX Homo sapiens.
 OS

PN W09402616-A.
 XX 03-FEB-1994.
 PD 20-JUL-1993; 93WO-US06703.
 PF 20-JUL-1992; 92US-0914281.
 XX (UNMI) UNIV MICHIGAN.
 PA (UNMI) UNIV MICHIGAN.
 XX Lowe JB;
 PI WPI: 1994-048874/06.
 DR N-PSDB: Q56907.
 XX DNA fragment encoding a glycosyltransferase - can be used for in
 PT vitro reactions to modify cell surface oligosaccharides) e.g
 PT blood gp. determinants, to protect against transplant rejection
 PT
 XX Disclosure; Fig 2; 249pp; English.
 PS The sequence is that of a human glycosyl transferase. The enzyme
 CC may be non glycosylated. This prevents premature loss of enzyme
 CC activity. It can also be used in in vitro reactions to modify cell
 CC surface oligosaccharide mols. e.g. blood group determinants.
 CC See also R45933-9.
 CC
 CC Sequence 394 AA;
 SQ

Query Match 90.3%; Score 28; DB 15; Length 394;
 Best Local Similarity 83.3%; Pred. No. 51;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNVGR 6
 |||||
 Db 36 mnvrgk 41

RESULT 13
 ID W13639 standard; Protein; 394 AA.
 AC W13639;
 XX 19-JUN-1997 (first entry)
 DT Murine alpha(1,3)-galactosyltransferase.
 DE Murine alpha(1,3)-galactosyltransferase.
 XX Alpha(1,3)-galactosyltransferase; glycosylation; oligosaccharide.
 KM Mus sp.
 OS Mus sp.
 XX W09709421-A1.
 PN 13-MAR-1997.
 PD 06-SEP-1996; 96WO-US13816.
 PF 08-SEP-1995; 95US-0525058.
 PR (UNMI) UNIV MICHIGAN.
 PA Legault DJ, Lowe JB;
 PI WPI: 1997-192897/17.
 DR N-PSDB: T61676.
 DR New recombinant fucosyltransferase proteins - useful for modifying
 PT cell surface oligosaccharide structures
 PT
 XX Example 2; Page 272-274; 329pp; English.
 PS

CC Murine UDP-Gal:beta-D-Gal(1,4)-D-GlcNAc alpha(1,3)-
CC galactosyltransferase (W13639) catalyses the a transglycosylation
CC reaction between UDP-Gal and N-acetyllactosamine and is associated
CC with surface-localised expression of Gal(alpha1-3)Gal linkages.
CC Its amino acid sequence was deduced from a cDNA clone (T61676) F9
CC obit. By transfecting COS-1 cells with cDNA derived from mouse F9
CC leterocarcinoma cells, and screening the transfected cells for
CC surface-localised Gal(alpha1-3)Gal linkages. When expressed in
CC animal cell lines, the enzyme provides specific capabilities with
CC respect to post-translational modification of the oligosaccharides
CC of expressed proteins or lipids. The enzyme can also be used to
CC raise antibodies and to screen for inhibitor cpds.

SO Sequence 394 AA:

Query Match 90.3%; Score 28; DB 18; Length 394;
Best Local Similarity 83.3%; Pred. No. 51;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKG 6
:||||:
Db 36 mnvkg 41

RESULT 14

W27869 W27869 standard; Protein: 94 AA.

AC W27869;

DT 20-AUG-1998 (first entry)

DE Amino acid sequence of a hypothetical protein in AmyB 5' region (ORF1).

XX Staphylococcus aureus protein; ribozyme; antisense sequence; control;
KM Staphylococcal gene; regulatory element; bacterial gene expression;
KW vaccine; Staphylococcal infection; food poisoning; scaled skin syndrome;
XX toxic shock syndrome.

OS Staphylococcus aureus.

XX Key Location/Qualifiers

FT Misc-difference 5 /note- "not specified"

FT Misc-difference 6 /note- "Gly encoded by GGN"

FT Misc-difference 62 /note- "not specified"

PN WO9730070-A1.

XX 21-AUG-1997.

XX 19-FEB-1997; 97WO-US02318.

XX 20-FEB-1996; 96US-0011888.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX Black MT, Burnham MK, Hodgson JE, Knowles DJC, Nicholas RO;

XX PI Pratt JM, Reichard RW, Rosenberg M, Ward JM;

XX WPI: 1997-424969/39.

XX N-PSDB: T83831.

XX Novel polypeptide(s) from Staphylococcus aureus strain WCUH29 - used
PT to isolate antimicrobial compounds, and in vaccines against S.

XX aureus infection

XX Claim 6; Pages 321-322; 989pp; English.

XX The present sequence represents a Staphylococcus aureus protein, that,

CC based on homology with a Thermococcus bacter thermophilus protein,
CC is believed to be a hypothetical 35.6 kDa protein in AmyB 5' region
CC (ORF1). The DNA sequence was isolated from a library of clones of
CC S. aureus WCUH 29 in Escherichia coli. The DNA sequence can be used
CC in the construction of ribozymes and antisense sequences to control the
CC expression of Staphylococcal genes. The DNA sequence is also useful as
CC a source of regulatory elements for the control of bacterial gene
CC expression. The present protein may be used to produce vaccines to
CC enable a host to produce specific antibodies with antibacterial action.
CC These vaccines and antibodies would protect a host against invasion by
CC S. aureus, and conditions relating to Staphylococcal infection,
CC e.g. Staphylococcal food poisoning, scaled skin syndrome, and toxic
CC shock syndrome.

SO Sequence 94 AA:

Query Match 87.1%; Score 27; DB 18; Length 94;
Best Local Similarity 66.7%; Pred. No. 18;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKG 6
:||||:
Db 78 Inlkg 83

RESULT 15

W38713 W38713 standard; Protein: 165 AA.

AC W38713;

DT 10-NOV-1998 (first entry)

DE S. pneumoniae glutamyl endopeptidase precursor.

XX Streptococcus pneumoniae protein; genetic immunisation; antagonist;
KM immunological response; inoculation; antibody production; inhibitor;
KW T cell immune response; antimicrobial compound; bacterial adhesion;
KW extracellular matrix protein; protein-mediated cell invasion; wound;
XX pathogenesis.

OS Streptococcus pneumoniae.

XX WO9743303-A1.

XX 20-NOV-1997.

XX 14-MAY-1997; 97WO-US07950.

XX 14-MAY-1996; 96US-0017670.

XX (SMIK) SMITHKLINE BEECHAM CORP.

XX (SMIK) SMITHKLINE BEECHAM PLC.

XX Black MT, Hodgson JE, Knowles DJC, Nicholas RO;

XX Stodola RK;

XX WPI: 1998-008793/01.

XX N-PSDB: T98755.

XX Novel Streptococcus pneumoniae proteins and related DNA - useful for
PT diagnosing anti-microbial agents for treatment of bacterial

XX infections

XX Claim 12; Page 448; 483pp; English.

XX This sequence represents a Streptococcus pneumoniae protein that, based
CC on homology with a B. licheniformis protein, is a glutamyl endopeptidase
CC precursor, and is encoded by a DNA sequence of the invention.

XX The DNA sequences were isolated from Streptococcus pneumoniae strain
CC 010093 (NCIMB 40794). The Streptococcus pneumoniae proteins of the
CC invention can be used to identify compounds which interact with and

CC inhibit or activate the activity of the proteins. Antagonists can be
 CC used to treat diseases caused by S. pneumoniae proteins, through genetic
 CC immunisation. They can also be used to induce an immunological response
 CC in a mammal by inoculation with the S. pneumoniae proteins or delivery
 CC of the encoding nucleic acids in a vector adequate to produce antibody
 CC and/or T cell immune responses to protect the animal from disease. The
 CC proteins can also be used to identify antimicrobial compounds which are
 CC capable of inhibiting their bioactivity. In particular the proteins of
 CC the invention can be used to prevent adhesion of bacteria to mammalian
 CC extracellular matrix proteins on in-dwelling devices or in wounds, to
 CC block protein-mediated mammalian cell invasion, and to block the normal
 CC progression of pathogenesis in infections initiated other than by the
 CC implantation of in-dwelling devices or other surgical techniques.

XX
 SQ Sequence 165 AA;

Query Match 87.1%; Score 27; DB 19; Length 165;
 Best Local Similarity 83.3%; Pred. No. 34;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNVKGK 6
 |||:|
 Db 13 mnvegr 18

Search completed: December 14, 2000, 10:11:09
 Job time: 43 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 14, 2000, 10:10:26 : Search time 27.22 seconds
(without alignments)
3.695 Million cell updates/sec

Title: US-09-051-034a-11
Perfect score: 31
Sequence: 1 MNVGR 6

Scoring table: BIOSUM62
Gapop 10.0, Gapext 0.5

Searched: 164575 seqs, 16761186 residues

Total number of hits satisfying chosen parameters: 164575

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/prodata2/1aa/5A.COMB.pep:*
2: /cgn2_6/prodata2/1aa/5B.COMB.pep:*
3: /cgn2_6/prodata2/1aa/6.COMB.pep:*
4: /cgn2_6/prodata2/1aa/PCTUS.COMB.pep:*
5: /cgn2_6/prodata2/1aa/beckfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	100.0	371	2	US-08-378-617A-10
2	28	90.3	117	1	US-08-249-013-6
3	28	90.3	117	2	US-08-886-863-6
4	28	90.3	117	4	PCT-US95-06764-6
5	28	90.3	368	2	US-08-378-617A-11
6	28	90.3	371	2	US-08-378-617A-12
7	28	90.3	376	2	US-08-704-548-2
8	28	90.3	374	1	US-07-914-281-4
9	28	90.3	394	1	US-08-393-246-4
10	28	90.3	394	1	US-08-525-058A-4
11	28	90.3	394	4	US-08-696-731-4
12	28	90.3	394	4	PCT-US91-00899-9
13	26	83.9	142	2	US-08-860-174A-7
14	26	83.9	274	2	US-08-860-174A-12
15	26	83.9	282	2	US-08-860-174A-10
16	26	83.9	394	3	US-08-705-771-21
17	25	80.6	145	1	US-08-551-171-3
18	25	80.6	145	3	US-08-902-233-3
19	25	80.6	146	1	US-08-551-171-4
20	25	80.6	146	3	US-08-902-233-4
21	25	80.6	170	1	US-08-551-171-1
22	25	80.6	170	3	US-08-902-233-1
23	25	80.6	171	1	US-08-551-171-2
24	25	80.6	171	3	US-08-902-233-2
25	25	80.6	174	1	US-08-551-171-7
26	25	80.6	174	3	US-08-902-233-7
27	25	80.6	175	1	US-08-551-171-8
28	25	80.6	175	3	US-08-902-233-8

29	25	80.6	198	1	US-08-439-725A-14	Sequence 14, Appl
30	25	80.6	198	2	US-08-441-629-12	Sequence 12, Appl
31	25	80.6	198	1	US-08-867-471-14	Sequence 10, Appl
32	25	80.6	198	2	US-08-438-439C-10	Sequence 20, Appl
33	25	80.6	198	2	US-08-438-439C-20	Sequence 15, Appl
34	25	80.6	198	3	US-08-705-245-15	Sequence 15, Appl
35	25	80.6	198	3	US-08-718-904-15	Sequence 15, Appl
36	25	80.6	198	3	US-09-023-082A-14	Sequence 14, Appl
37	25	80.6	198	4	US-08-776-207-12	Sequence 12, Appl
38	25	80.6	198	4	PCT-US95-09172-12	Sequence 12, Appl
39	25	80.6	207	1	US-08-551-171-5	Sequence 5, Appl1
40	25	80.6	207	3	US-08-902-233-5	Sequence 5, Appl1
41	25	80.6	208	1	US-08-551-171-6	Sequence 6, Appl1
42	25	80.6	208	1	US-08-464-590A-16	Sequence 16, Appl
43	25	80.6	208	1	US-08-462-169B-14	Sequence 14, Appl
44	25	80.6	208	2	US-08-207-412B-11	Sequence 11, Appl
45	25	80.6	208	2	US-08-951-822-27	Sequence 27, Appl

ALIGNMENTS

RESULT 1
US-08-378-617A-10
Sequence 10, Application US/08378617A
Patent No. 5849991
GENERAL INFORMATION:
APPLICANT: d'Apice, Anthony J.F.
APPLICANT: Rodins, Martin J.
APPLICANT: Rodins, Allan J.
APPLICANT: Crawford, Robert J.
APPLICANT: Ratchjen, Peter D.
TITLE OF INVENTION: MATERIALS AND METHODS FOR MANAGEMENT OF
TITLE OF INVENTION: HYPERACUTE REJECTION IN HUMAN XENOTRANSPLANTATION
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 120 South Sixth Street, Suite 2500
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/378,617A
FILING DATE: 26-JAN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Ellinger, Mark S.
REGISTRATION NUMBER: 34,812
REFERENCE/DOCKET NUMBER: 06868/005001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (612) 335-5070
TELEFAX: (612) 288-9696
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 371 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-378-617A-10

Query Match 100.0%; Score 31; DB 2; Length 371;
Best Local Similarity 100.0%; Pred. No. 12;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 MNVGR 6
|||||

Db 1 MNVGR 6

RESULT 2

US-08-249-013-6
Sequence 6, Application US/08249013
Patent No. 5643754

GENERAL INFORMATION:

APPLICANT: Haake, David A.

TITLE OF INVENTION: CLONED leptospira OUTER MEMBRANE PROTEIN

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Spensley Horn Jubas & Lubitz

STREET: 1880 Century Park East, Suite 500

CITY: Los Angeles

STATE: California

COUNTRY: USA

ZIP: 90067

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/249,013

FILING DATE: 25-MAY-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Tumarkin Ph.D., Lisa A.,

REGISTRATION NUMBER: P-38,347

REFERENCE/DOCKET NUMBER: PD-3602

TELEPHONE: (619) 455-5110

TELEFAX: (619) 455-5110

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 117 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

IMMEDIATE SOURCE:

CLONE: lula

FEATURE:

NAME/KEY: Protein

LOCATION: 1..117

US-08-249-013-6

Query Match 90.3%; Score 28; DB 1; Length 117;

Best Local Similarity 83.3%; Pred. No. 17;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1 MNVGR 6

Db 25 MNVGR 30

RESULT 3

US-08-886-863-6

Sequence 6, Application US/0886863

Patent No. 5824321

GENERAL INFORMATION:

APPLICANT: Haake, David A.

TITLE OF INVENTION: CLONED leptospira OUTER MEMBRANE PROTEIN

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Spensley Horn Jubas & Lubitz

STREET: 1880 Century Park East, Suite 500

CITY: Los Angeles

STATE: California

COUNTRY: USA

ZIP: 90067

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/886,863

FILING DATE: 01-JUL-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/249,013

FILING DATE: 25-MAY-1994

ATTORNEY/AGENT INFORMATION:

NAME: Tumarkin Ph.D., Lisa A.,

REGISTRATION NUMBER: P-38,347

REFERENCE/DOCKET NUMBER: PD-3602

TELEPHONE: (619) 455-5110

TELEFAX: (619) 455-5110

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 117 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

IMMEDIATE SOURCE:

CLONE: lula

FEATURE:

NAME/KEY: Protein

LOCATION: 1..117

US-08-886-863-6

Query Match 90.3%; Score 28; DB 2; Length 117;

Best Local Similarity 83.3%; Pred. No. 17;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Db 1 MNVGR 6

Db 25 MNVGR 30

RESULT 4

PCT-US95-06764-6

Sequence 6, Application PC/TUS9506764

GENERAL INFORMATION:

APPLICANT: The Regents of the University of California

TITLE OF INVENTION: CLONED leptospira OUTER MEMBRANE PROTEIN

NUMBER OF SEQUENCES: 10

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson

STREET: 4225 Executive Square, Suite 1400

CITY: La Jolla

STATE: California

COUNTRY: USA

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/06764

FILING DATE: 25-MAY-1995

CLASSIFICATION:

ATTORNEY/AGENT INFORMATION:

NAME: Haile, Ph.D., Lisa A.,

REGISTRATION NUMBER: 38,347

REFERENCE/DOCKET NUMBER: FD3602

TELEPHONE: (619) 678-5070

TELEFAX: (619) 678-5099

;; INFORMATION FOR SEQ ID NO: 6:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 117 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; IMMEDIATE SOURCE:
;; CLONE: Iuta
;; FEATURE:
;; NAME/KEY: Protein
;; LOCATION: 1..117
PCT-US95-06764-6

Query Match 90.3% Score 28; DB 4; Length 117;
Best Local Similarity 83.3% Pred. No. 17;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNVGR 6
|||:
Db 25 MNVGR 30

RESULT 5
US-08-378-617A-11
; Sequence 11, Application US/08378617A.
; Patent No. 5849991
; GENERAL INFORMATION:
; APPLICANT: d'Apice, Anthony J.F.
; APPLICANT: Pearse, Martin J.
; APPLICANT: Robins, Allan J.
; APPLICANT: Crawford, Robert J.
; APPLICANT: Rathjen, Peter D.
; TITLE OF INVENTION: MATERIALS AND METHODS FOR MANAGEMENT OF
; TITLE OF INVENTION: HYPERACUTE REJECTION IN HUMAN XENOTRANSPLANTATION
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 120 South Sixth Street, Suite 2500
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/378,617A
; FILING DATE: 26-JAN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ellinger, Mark S.
; REGISTRATION NUMBER: 34,812
; REFERENCE/DOCKET NUMBER: 06868/005001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (612) 335-5070
; TELEFAX: (612) 288-9696
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-378-617A-11

Query Match 90.3% Score 28; DB 2; Length 368;
Best Local Similarity 83.3% Pred. No. 56;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNVGR 6
|||:
Db .1 MNVGR 6

RESULT 6
US-08-378-617A-12
; Sequence 12, Application US/08378617A
; Patent No. 5849991
; GENERAL INFORMATION:
; APPLICANT: d'Apice, Anthony J.F.
; APPLICANT: Pearse, Martin J.
; APPLICANT: Robins, Allan J.
; APPLICANT: Crawford, Robert J.
; APPLICANT: Rathjen, Peter D.
; TITLE OF INVENTION: MATERIALS AND METHODS FOR MANAGEMENT OF
; TITLE OF INVENTION: HYPERACUTE REJECTION IN HUMAN XENOTRANSPLANTATION
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 120 South Sixth Street, Suite 2500
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/378,617A
; FILING DATE: 26-JAN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ellinger, Mark S.
; REGISTRATION NUMBER: 34,812
; REFERENCE/DOCKET NUMBER: 06868/005001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (612) 335-5070
; TELEFAX: (612) 288-9696
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 371 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-378-617A-12

Query Match 90.3% Score 28; DB 2; Length 371;
Best Local Similarity 83.3% Pred. No. 56;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNVGR 6
|||:
Db 1 MNVGR 6

RESULT 7
US-08-704-548-2
; Sequence 2, Application US/08704548
; Patent No. 5879675
; GENERAL INFORMATION:
; APPLICANT: GALILI, URI
; APPLICANT: REPICK, PATRICIA M.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR VACCINES
; TITLE OF INVENTION: COMPRISING ALPHA-GALACTOSYL EPITOPES
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEIDEL, GONDA, LAVORGNA & MONACO, P.C.
; STREET: Suite 1800, Two Penn Center Plaza
; CITY: Philadelphia

STATE: PA
COUNTRY: USA
ZIP: 19102
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/704,548
FILING DATE: 11-SEP-1996
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Monaco, Daniel A.
REGISTRATION NUMBER: 30,480
REFERENCE/DOCKET NUMBER: 8760-2 C11
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-8383
TELEFAX: (215) 568-5549
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 376 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-704-548-2

Query Match 90.3%; Score 28; DB 2; Length 376;
Best Local Similarity 83.3%; Pred. No. 57;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNVGR 6
Db 1 MNVGR 6

RESULT 8
US-07-914-281-4
Sequence 4, Application US/07914281
Patent No. 5324663
GENERAL INFORMATION:
APPLICANT: LOWE, JOHN B.
TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESS: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESS: P.C.
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/914,281
FILING DATE: 19920720
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Lavalleye, Jean-Paul M. P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-060-55
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)486-2347
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 394 amino acids
TYPE: AMINO ACID
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-07-914-281-4

Query Match 90.3%; Score 28; DB 1; Length 394;
Best Local Similarity 83.3%; Pred. No. 60;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNVGR 6
Db 36 MNVGR 41

RESULT 9
US-08-393-246-4
Sequence 4, Application US/08393246
Patent No. 5595900
GENERAL INFORMATION:
APPLICANT: LOWE, JOHN B.
TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESS: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESS: P.C.
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/393,246
FILING DATE: 30-MAR-1994
APPLICATION NUMBER: US 08/220,433
FILING DATE: 20-JUL-1992
APPLICATION NUMBER: US 07/914,281
ATTORNEY/AGENT INFORMATION:
NAME: Lavalleye, Jean-Paul M. P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-060-55
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 394 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-393-246-4

Query Match 90.3%; Score 28; DB 1; Length 394;
Best Local Similarity 83.3%; Pred. No. 60;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNVGR 6

DB 36 MNVKGK 41

RESULT 10
US-08-525-058A-4

Sequence 4, Application US/08525058A
Patent No. 5770420

GENERAL INFORMATION:

APPLICANT: LOWE, JOHN B.

TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS

TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,

TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION

TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES

NUMBER OF SEQUENCES: 23

CORRESPONDENCE ADDRESS: 23

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT, P.C.

STREET: 1755 Jefferson Davis Highway, Fourth Floor

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/525.058A

FILING DATE:

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Lavalleye, Jean-Paul M. P.

REGISTRATION NUMBER: 31,451

REFERENCE/DOCKET NUMBER: 2363-060-55

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)521-4500

TELEFAX: (703)486-2347

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 394 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-525-058A-4

Query Match

Best Local Similarity 90.3%; Score 28; DB 1; Length 394;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DB 36 MNVKGK 41

RESULT 11

US-08-696-731-4

Sequence 4, Application US/08696731

Patent No. 5955347

GENERAL INFORMATION:

APPLICANT: LOWE, JOHN B.

TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS

TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,

TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION

TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS: 23

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT, P.C.

STREET: 1755 Jefferson Davis Highway, Fourth Floor

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/696.731

FILING DATE: 14-AUG-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Lavalleye, Jean-Paul M. P.

REGISTRATION NUMBER: 31,451

REFERENCE/DOCKET NUMBER: 2363-060-55

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)521-4500

TELEFAX: (703)486-2347

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 394 amino acids

TYPE: amino acid

TOPOLOGY: unknown

MOLECULE TYPE: protein

US-08-696-731-4

Query Match

Best Local Similarity 90.3%; Score 28; DB 2; Length 394;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

DB 36 MNVKGK 41

RESULT 12

PCT-US91-00899-9

Sequence 9, Application PC/TUS9100899

GENERAL INFORMATION:

APPLICANT: LOWE, John B.

TITLE OF INVENTION: Method and Products for the Synthesis of

TITLE OF INVENTION: Oligosaccharide Structures on Glycoproteins, Glycolipids,

TITLE OF INVENTION: or as Free Molecules, and for the Isolation of Cloned

TITLE OF INVENTION: Genetic Sequences That Determine These Structures

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS: 23

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MATER & NEUSTADT, P.C.

STREET: 1755 Jefferson Davis Highway, Suite 400

CITY: Arlington

STATE: Virginia

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US91/00899

FILING DATE: 19910214

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Lavalleye Ph.D., Jean-Paul

REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-021-55 PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-5940
TELEFAX: (703)486-2347
FTELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 394 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE:
TISSUE TYPE: Blood
PCT-US91-00899-9

Query Match 90.3%; Score 28; DB 4; Length 394;
Best Local Similarity 83.3%; Pred. No. 60;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNVGR 6
DB 86 MNVGR 41

RESULT 13
US-08-860-174A-7
Sequence 7, Application US/08860174A
Patent No. 5989830
GENERAL INFORMATION:
APPLICANT: DAVIS, Paul James
APPLICANT: VAN DER LOGT, Cornelis Paul Erik
APPLICANT: VERHOEIJEN, Martine Elisa
APPLICANT: WILSON, Steve
TITLE OF INVENTION: A BIFUNCTIONAL OR BIVALENT
TITLE OF INVENTION: ANTIBODY FRAGMENT ANALOGUE
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESSES:
ADDRESSEE: PILLSBURY, MADISON & SUTRO LLP
STREET: 1100 New York Avenue, N.W.
STREET: 9th Floor, East Tower
CITY: WASHINGTON, D.C.
STATE:
COUNTRY: UNITED STATES
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS DOS Text
SOFTWARE: MS WORD, WINDOWS 95, Ver. 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/860,174A
FILING DATE: June 16, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95307332.7
FILING DATE: October 16, 1995
APPLICATION NUMBER: PCT/EP/96/03605
FILING DATE: August 14, 1996
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 142 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-860-174A-7

Query Match 83.9%; Score 26; DB 2; Length 142;
Best Local Similarity 100.0%; Pred. No. 58;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 NVKGR 6
DB 85 NVKGR 89

RESULT 14
US-08-860-174A-12
Sequence 12, Application US/08860174A
Patent No. 5989830
GENERAL INFORMATION:
APPLICANT: DAVIS, Paul James
APPLICANT: VAN DER LOGT, Cornelis Paul Erik
APPLICANT: VERHOEIJEN, Martine Elisa
APPLICANT: WILSON, Steve
TITLE OF INVENTION: A BIFUNCTIONAL OR BIVALENT
TITLE OF INVENTION: ANTIBODY FRAGMENT ANALOGUE
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESSES:
ADDRESSEE: PILLSBURY, MADISON & SUTRO LLP
STREET: 1100 New York Avenue, N.W.
STREET: 9th Floor, East Tower
CITY: WASHINGTON, D.C.
STATE:
COUNTRY: UNITED STATES
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS DOS Text
SOFTWARE: MS WORD, WINDOWS 95, Ver. 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/860,174A
FILING DATE: June 16, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95307332.7
FILING DATE: October 16, 1995
APPLICATION NUMBER: PCT/EP/96/03605
FILING DATE: August 14, 1996
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 274 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-860-174A-12

Query Match 83.9%; Score 26; DB 2; Length 274;
Best Local Similarity 100.0%; Pred. No. 11e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 NVKGR 6
DB 85 NVKGR 89

RESULT 15
US-08-860-174A-10
Sequence 10, Application US/08860174A
Patent No. 5989830
GENERAL INFORMATION:
APPLICANT: DAVIS, Paul James
APPLICANT: VAN DER LOGT, Cornelis Paul Erik
APPLICANT: VERHOEIJEN, Martine Elisa
APPLICANT: WILSON, Steve
TITLE OF INVENTION: A BIFUNCTIONAL OR BIVALENT
TITLE OF INVENTION: ANTIBODY FRAGMENT ANALOGUE
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESSES:
ADDRESSEE: PILLSBURY, MADISON & SUTRO LLP
STREET: 1100 New York Avenue, N.W.
STREET: 9th Floor, East Tower

CITY: WASHINGTON, D.C.
STATE:
COUNTRY: UNITED STATES
ZIP: 20005-3918
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS DOS Text
SOFTWARE: MS WORD, WINDOWS 95, Ver. 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/860,174A
FILING DATE: June 16, 1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: EP 95307332.7
FILING DATE: October 16, 1995
APPLICATION NUMBER: PCT/EP/96/03605
FILING DATE: August 14, 1996
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 282 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-860-174A-10

Query Match 83.9%: Score 26; DB 2; Length 282;
Best Local Similarity 100.0%; Pred. No. 1.2e+02;
Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 NVKGR 6
| | | | |
DB 85 NVKGR 89

Search completed: December 14, 2000, 10:11:39
Job time: 73 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 14, 2000, 10:10:27 ; Search time 27.57 Seconds
(without alignments)
13.811 Million cell updates/sec

Title: US-09-051-034a-11

Perfect score: 31
Sequence: 1 MNVGR 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters: 182106

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	100.0	371	2 I46583	alpha-1,3-galactos
2	31	100.0	378	2 T30965	hypothetical prote
3	30	96.8	454	2 S61052	hypothetical prote
4	28	90.3	26	2 A39769	N-acetylglucosamin
5	28	90.3	308	2 F70917	probable integral m
6	28	90.3	368	2 A44785	N-acetylglucosamin
7	28	90.3	371	2 I49698	alpha-1,3-galactos
8	28	90.3	376	2 A56460	N-acetylglucosamin
9	28	90.3	394	2 A34417	alpha-1,3-mannosyl
10	28	90.3	473	2 E75097	hypothetical prote
11	28	90.3	725	2 S01042	aerobactin recepto
12	27	87.1	287	2 C71207	probable ribose-ph
13	27	87.1	322	1 B64347	hypothetical prote
14	27	87.1	334	1 A69011	conserved hypotet
15	27	87.1	386	1 BLEC	leucine/isoleucine
16	27	87.1	420	2 C69532	conserved hypotet
17	27	87.1	559	1 WMBE51	71K upper matrix p
18	27	87.1	686	2 T23721	hypothetical prote
19	27	87.1	763	2 T21006	hypothetical prote
20	27	87.1	1160	2 T27487	hypothetical prote
21	27	87.1	1160	2 B47648	perlecan homolog u
22	27	87.1	2277	2 B47648	perlecan homolog u
23	27	87.1	2295	2 T19820	hypothetical prote
24	27	87.1	2482	2 T19819	hypothetical prote
25	27	87.1	2482	2 A47648	perlecan homolog u
26	27	87.1	3375	2 T19821	hypothetical prote
27	26	83.9	119	2 T00151	hypothetical prote
28	26	83.9	149	2 A69948	phage-related prot
29	26	83.9	177	2 S39488	mannose-binding le

30	26	83.9	184	2 S12260	hypothetical prote
31	26	83.9	235	2 S33448	resiniferatoxin-bl
32	26	83.9	254	2 C70387	dehydrogenase - Aq
33	26	83.9	259	2 E72861	AcOrf-92 protein -
34	26	83.9	259	2 T41832	ACMNPV orf92 - Bom
35	26	83.9	278	2 H75259	epoxide hydrolase-
36	26	83.9	282	2 T10362	hypothetical prote
37	26	83.9	348	2 C70415	cation efflux syst
38	26	83.9	375	2 A64398	hypothetical prote
39	26	83.9	405	2 T34119	hypothetical prote
40	26	83.9	444	2 T05346	ferredoxin--NADP+
41	26	83.9	466	2 H64904	hypothetical prote
42	26	83.9	506	2 T02400	beta-glucosidase h
43	26	83.9	512	2 T05107	hypothetical prote
44	26	83.9	560	1 VHIVBC	nucleoprotein - In
45	26	83.9	560	1 VHIVBM	nucleoprotein - In

ALIGNMENTS

RESULT 1
I46583
alpha-1,3-galactosyltransferase - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 21-Feb-1997 #sequence_revision 21-Feb-1997 #text_change 29-Sep-1999
C:Accession: I46583
R:Strahan, K.M.; Gu, F.; Preece, A.F.; Gustavsson, I.; Andersson, L.; Gustafsson, K.
Immunogenetics 41, 101-105, 1995
A:Title: CDNA sequence and chromosome localization of pig alpha 1,3 galactosyltransfe
A:Reference number: I46583; MUID:95104914
A:Accession: I46583
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-371 <STR>
A:Cross-references: GB:U36152; NID:9642635; PIDN:AAA73558.1; PID:9642636
C:Genetics:
A:Gene: GCFAL
C:Superfamily: histo-blood group 1 transferase

Query Match 100.0%; Score 31; DB 2; Length 371;
Best Local Similarity 100.0%; Pred. No. 11;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNVGR 6
|||||
DB 1 MNVGR 6

RESULT 2
T30965
hypothetical protein R04B3.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jan-2000
C:Accession: T30965
R:Gattung, S.; Pauley, A.
Submitted to the EMBL Data Library, August 1999
A:Description: The sequence of C. elegans cosmid R04B3.
A:Reference number: Z20947
A:Accession: T30965
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-378 <GAT>
A:Cross-references: EMBL:U50198; PIDN:AAA91360.1
A:Experimental source: strain Bristol N2
C:Genetics:
A:Map position: X
A:Introns: 42/1; 182/3; 222/2; 251/3; 322/1
A:Note: R04B3.2
C:Superfamily: N4-(beta-N-acetylglucosaminyl)-L-asparaginase

Query Match 100.0%; Score 31; DB 2; Length 378;
 Best Local Similarity 100.0%; Pred. No. 11;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNVKGK 6
 DB 327 MNVKGK 332

RESULT 3
 561052
 hypothetical protein YDL161w - yeast (Saccharomyces cerevisiae)
 N:Allegiate names: hypothetical protein D1515
 C:Species: Saccharomyces cerevisiae
 C:Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 29-Oct-1999
 C:Accession: S61052; S67713
 R:Pol: T.M.
 Submitted to the EMBL Data Library, November 1995
 A:Reference number: S61010
 A:Accession: S61052
 A:Molecule type: DNA
 A:Residues: 1-454 <POB>
 A:Cross-references: EMBL:267750; NID:q1061256; PIDN:CAA91585.1; PID:q1061278
 R:Pol: T.M.
 Submitted to the Protein Sequence Database, July 1996
 A:Reference number: S67708
 A:Accession: S67713
 A:Molecule type: DNA
 A:Residues: 1-454 <POB>
 A:Cross-references: EMBL:274210; NID:q1431255; PIDN:CAA98736.1; PID:q143125
 A:Experimental source: strain S288C
 C:Genetics:
 A:Map position: 4L

Query Match 96.8%; Score 30; DB 2; Length 454;
 Best Local Similarity 83.3%; Pred. No. 23;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNVKGK 6
 DB 245 MNVKGK 150

RESULT 4
 A39769
 N-acetylglucosaminide alpha-1,3-galactosyltransferase (EC 2.4.1.151) homolog - human (fr
 C:Species: Homo sapiens (man)
 C:Date: 14-Feb-1992 #sequence_revision 14-Feb-1992 #text_change 13-Sep-1998
 C:Accession: A39769
 R:Joziasse, D.H.; Shaper, J.H.; Jabs, E.W.; Shaper, N.L.
 J. Biol. Chem. 266: 6891-6998, 1991
 A:Title: Characterization of an alpha1-->3-galactosyltransferase homologue on human chro
 A:Reference number: A39769; MUID:91201351
 A:Accession: A39769
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-26 <Y0Z>
 C:Superfamily: histo-blood group 1 transferase
 C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 90.3%; Score 28; DB 2; Length 26;
 Best Local Similarity 83.3%; Pred. No. 4.4;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNVKGK 6
 DB 1 MNVKGK 6

RESULT 5
 F70917

probable integral membrane protein - Mycobacterium tuberculosis (strain H37RV)
 C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 20-Jun-2000
 C:Accession: F70917
 R:Cole, S.T.; Broesch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon
 ; Connor, R.; Davies, R.; Devlin, K.; Feltwell, T.; Gentles, S.; Hamlin, N.; Holroyd,
 Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellon, S.; Squares, S.
 Nature 393, 537-544, 1998
 A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.
 A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno
 A:Reference number: A70500; MUID:98295987
 A:Accession: F70917
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-308 <COL>
 A:Cross-references: GB:295844; GB:AL123456; NID:g3250713; PIDN:CAB09270.1; PID:g21310
 A:Experimental source: strain H37RV
 C:Genetics:
 A:Gene: clab
 C:Superfamily: heme O synthase

Query Match 90.3%; Score 28; DB 2; Length 308;
 Best Local Similarity 83.3%; Pred. No. 46;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNVKGK 6
 DB 1 MNVKGK 6

RESULT 6
 A44785
 N-acetylglucosamine 3-alpha-galactosyltransferase (EC 2.4.1.124) - bovine
 C:Species: Bos primigenius taurus (cattle)
 C:Date: 19-Mar-1993 #sequence_revision 19-Mar-1993 #text_change 29-Sep-1999
 C:Accession: A44785
 R:Joziasse, D.H.; Shaper, J.H.; Van den Bijnden, D.H.; Van Tunen, A.J.; Shaper, N.L.
 J. Biol. Chem. 264, 14290-14297, 1989
 A:Title: Bovine alpha1-->3-galactosyltransferase: isolation and characterization of a
 A:Reference number: A44785; MUID:89340543
 A:Accession: A44785
 A:Status: not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-368 <Y0Z>
 A:Cross-references: GB:J04989; NID:g163123; PIDN:AAA30558.1; PID:g163124
 C:Superfamily: histo-blood group 1 transferase
 C:Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein

Query Match 90.3%; Score 28; DB 2; Length 368;
 Best Local Similarity 83.3%; Pred. No. 55;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNVKGK 6
 DB 1 MNVKGK 6

RESULT 7
 I49698
 alpha-1,3-galactosyltransferase - mouse
 C:Species: Mus musculus (house mouse)
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 29-Sep-1999
 C:Accession: I49698
 R:Joziasse, D.H.; Shaper, N.L.; Kim, D.; van der Eijnden, D.H.; Shaper, J.H.
 J. Biol. Chem. 267, 5534-5541, 1992
 A:Title: Murine alpha-1,3-galactosyltransferase: A single gene locus specifies four i
 A:Reference number: A42117; MUID:92184813
 A:Accession: I49698
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-371 <RES>

A:Cross-references: GB:M85153; NID:q193563; PIDN:AAA37711.1; PID:q457142
A:Genetics:
A:Gene: Gata-1
C:Superfamily: histo-blood group 1 transferase

Query Match 90.3%; Score 28; DB 2; Length 371;
Best Local Similarity 83.3%; Pred. No. 55;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVGR 6
|||||
DB 1 MNVGR 6

RESULT 8
A56480
N-acetylglucosaminidase alpha-1,3-galactosyltransferase (EC 2.4.1.151) - marmoset

C:Species: Callithrix sp.
C>Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 29-Sep-1999
C:Accession: A56480

R:Henion, T. R.; Machner, B. A.; Anaraki, F.; Gallili, U.

Glycobiology 4, 193-201, 1994

A:Title: Defining the minimal size of catalytically active primate alpha1,3 galactosyltr

A:Reference number: A56480; MUID:94331837

A:Accession: A56480

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-376 <HEN>

A:Cross-references: GB:S71333; NID:q558051; PIDN:AAB31587.1; PID:q558052

A>Note: Authors translated the codon GTC for residue 251 as Ser

C:Superfamily: histo-blood group 1 transferase

C:Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein

Query Match 90.3%; Score 28; DB 2; Length 376;
Best Local Similarity 83.3%; Pred. No. 56;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVGR 6
|||||
DB 1 MNVGR 6

RESULT 9

A34417
alpha-1,3-mannosyl-glycoprotein beta-1,4-N-acetylglucosaminyltransferase (EC 2.4.1.145)

C:Species: Mus musculus (house mouse)

C>Date: 15-Jun-1990 #sequence_revision 15-Jun-1990 #text_change 29-Sep-1999

C:Accession: A34417

R:Larsen, R. D.; Rajan, V. P.; Ruff, M. M.; Kukowska-Latallo, J.; Cummings, R. D.; Lowe, J. E

Proc. Natl. Acad. Sci. U.S.A. 86, 8227-8231, 1989

A:Title: Isolation of a cDNA encoding a murine UDPgalactose:beta-D-galactosyl-1,4-N-acet

A:Reference number: A34417; MUID:90046769

A:Accession: A34417

A>Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-394 <LAR>

A:Cross-references: GB:M26925; NID:q193419; PIDN:AAA37657.1; PID:q309242

C:Superfamily: histo-blood group 1 transferase

C:Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein

Query Match 90.3%; Score 28; DB 2; Length 394;
Best Local Similarity 83.3%; Pred. No. 59;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVGR 6
|||||
DB 36 MNVGR 41

RESULT 10

E75097
hypothetical protein PAB0780 - Pyrococcus abyssi (strain Orsay)

C:Species: Pyrococcus abyssi

C>Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000

C:Accession: E75097

R:anonymous, Genoscope

submitted to the EMBL Data Library, July 1999

A:Description: Pyrococcus abyssi genome sequence: Insights into archaeal chromosome s

A:Reference number: A75001

A:Accession: E75097

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-473 <KAN>

A:Cross-references: GB:AJ248286; GB:AL096836; NID:q5458366; PIDN:CAB50082.1; PID:q545

A:Experimental source: strain Orsay

C:Genetics:
A:Gene: PAB0780

Query Match 90.3%; Score 28; DB 2; Length 473;
Best Local Similarity 83.3%; Pred. No. 70;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVGR 6
|||||
DB 331 MNVGR 336

RESULT 11

S01042
aerobactin receptor precursor lntA [validated] - Escherichia coli plasmid COLV-K30

N:Alternate names: cloacin receptor

C:Species: Escherichia coli

C>Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 26-May-2000

C:Accession: S01042

R:Krone, W. U. A.; Stegehuis, F.; Koningsstein, G.; van Doorn, C.; Roosendaal, B.; de Gr

FEMS Microbiol. Lett. 26, 153-161, 1985

A:Title: Characterization of the pCOLV-K30 encoded cloacin Df13/aerobactin outer memb

e sequence and primary structure.

A:Reference number: S01042

A:Accession: S01042

A:Molecule type: DNA

A:Residues: 1-725 <KRO>

A:Cross-references: EMBL:X05874

C:Genetics:
A:Gene: lntA

A:Genome: plasmid COLV-K30

C:Function:
A:Description: functions as outer membrane receptor for ferric aerobactin [validated];

F:1-25/Domain: signal sequence #status predicted <SIG>

F:26-725/Product: cloacin receptor #status predicted <MAT>

Query Match 90.3%; Score 28; DB 2; Length 725;
Best Local Similarity 83.3%; Pred. No. 1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVGR 6
|||||
DB 92 MNVGR 97

RESULT 12

C71207
probable ribose-phosphate pyrophosphokinase - Pyrococcus horikoshii

C:Species: Pyrococcus horikoshii

C>Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000

C:Accession: C71207

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Se

M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu

DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophil

A:Reference number: A71000; MUID:98344137

A:Accession: C71207
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-287 <RAM>
 A:Cross-references: GB:AP000007; NID:g3236134; PID:g3258367
 A:Experimental source: strain OT3
 A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
 C:Genetics:
 A:Gene: PH1923
 C:Superfamily: ribose-phosphate pyrophosphokinase catalytic chain

Query Match 87.1%; Score 27; DB 2; Length 287;
 Best Local Similarity 83.3%; Pred. No. 75;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MNVGR 6
 Db 200 VNVGR 205

RESULT 13

B64347
 hypothetical protein MJ0378 - Methanococcus jannaschii
 C:Species: Methanococcus jannaschii
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: B64347
 R:Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Retch, C.I.; Overbeek, R.; Kirkness, E.F.; Weissbrock, K.G.; Merrick, J.M.; Glodek, A.; Ison, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
 Science 273, 1058-1073, 1996
 A:Authors: Kahne, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
 A:Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii
 A:Reference number: A64300; MUID:96337999
 A:Accession: B64347
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-322 <BUL>
 A:Cross-references: GB:U67490; GB:L77117; NID:g1591074; PID:g1499166; TIGR:MJ0378
 C:Genetics:
 A:Map position: FOR343921-344889
 C:Superfamily: conserved hypothetical protein MJ0378

Query Match 87.1%; Score 27; DB 1; Length 322;
 Best Local Similarity 83.3%; Pred. No. 83;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MNVGR 6
 Db 146 MNVGR 151

RESULT 14

A69011
 conserved hypothetical protein MTH1084 - Methanobacterium thermoautotrophicum (strain De C:Species: Methanobacterium thermoautotrophicum
 C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
 C:Accession: A69011
 R:Smith, D.R.; Doucette-Stamm, L.A.; Deloughery, C.; Lee, H.; Dubois, J.; Aldredge, T.; Qu, D.; Spadator, R.; Vicaltre, R.; Wang, Y.; Wierzbowski, J.; Gibson, R.; Jiwani, N.; K. S. Church, G.M.; Daniels, C.J.; Mao, J.; Rice, P.; Noelling, J.; Reeve, J.N.
 J. Bacteriol. 179, 7135-7159, 1997
 A:Title: Complete genome sequence of Methanobacterium thermoautotrophicum Delta H: functional genome number: A69000; MUID:98037514
 A:Reference number: A69000; MUID:98037514
 A:Accession: A69011
 A>Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-334 <MTH>
 A:Cross-references: GB:AE000879; GB:AE000666; NID:g2622175; PID:g2622184
 A:Experimental source: strain Delta H
 C:Genetics:
 A:Gene: MTH1084

A:Start codon: TTG
 C:Superfamily: conserved hypothetical protein MJ0378

Query Match 87.1%; Score 27; DB 1; Length 334;
 Best Local Similarity 83.3%; Pred. No. 86;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MNVGR 6
 Db 158 MNVGR 163

RESULT 15

BLEC
 leucine/isoleucine/valine-binding protein precursor - Escherichia coli
 N:Alternate names: leucine transport protein livJ precursor; LIV-binding protein
 C:Species: Escherichia coli
 C>Date: 31-May-1979 #sequence_revision 05-Dec-1997 #text_change 16-Jul-1999
 C:Accession: G65142; A23576; A37074; A03415; S47679; I55524
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.
 A:Reference number: A64720; MUID:97426517
 A:Accession: G65142
 A>Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-386 <BLAT>
 A:Cross-references: GB:AE000422; GB:U00096; NID:g1789868; PIDN:AACT6485.1; PID:g17898
 A:Experimental source: strain K-12, Substrain MG155
 R:Landick, R.; Oxender, D.L.
 J. Biol. Chem. 260, 8257-8261, 1985
 A:Title: The complete nucleotide sequences of the Escherichia coli LIV-BP and LS-BP g
 A:Reference number: A94677; MUID:85234531
 A:Accession: A23576
 A:Molecule type: DNA
 A:Residues: 20-21, 'T', '23-30, 'L', '32-88, 'A', '90-386 <LAN>
 A:Cross-references: GB:J05516; GB:K02178; GB:M10426; GB:M13166; NID:g14663
 R:Adams, M.D.; Wagner, L.M.; Graddis, T.J.; Landick, R.; Antonucci, T.K.; Gibson, A.L.
 J. Biol. Chem. 265, 11436-11443, 1990
 A:Title: Nucleotide sequence and genetic characterization reveal six essential genes
 A:Reference number: A37074; MUID:90307651
 A:Accession: A37074
 A:Molecule type: DNA
 A:Residues: 20-21, 'T', '23-30, 'L', '32-88, 'A', '90-386 <AD>
 A:Cross-references: GB:J05516; NID:g146630; PIDN:AA83881.1; PID:g146631
 R:Orchinnikov, Y.A.; Aidanova, N.A.; Grinkevich, V.A.; Arzamazova, N.M.; Moroz, I.N.
 FEBS Lett. 78, 313-316, 1977
 A:Title: The primary structure of a Leu, Ile and Val (LIV)-binding protein from Esche
 A:Reference number: A03415; MUID:77225636
 A:Accession: A03415
 A:Molecule type: protein
 A:Residues: 43-304, 'AD', '307-383, 'AD', '386 <OVC>
 A:Experimental source: strain K12
 A:Note: this protein was isolated from the periplasmic space of the cell
 R:Plunkett, G.
 submitted to the EMBL Data Library, March 1994
 A:Reference number: S47666
 A:Accession: S47679
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-386 <PLU>
 A:Cross-references: EMBL:U00039; NID:g466582; PIDN:AA81435.1; PID:g912457
 A:Note: this reported protein translation starts with a GTG codon; it has an extra 23
 R:Antonucci, T.K.; Landick, R.; Oxender, D.L.
 J. Cell. Biochem. 29, 209-216, 1985
 A:Title: The leucine binding proteins of Escherichia coli as models for studying the
 A:Reference number: I55524; MUID:86086153
 A:Accession: I55524
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 20-21, 'T', '23-30, 'L', '32-88, 'A', '90-386 <RES>

A:Cross-references: GB:M29377; MID:G146628; PIDN:AAA24075.1; PID:G146629
 C:Gene: livj
 A:Map position: 76 min
 C:Function: specifically binds Leu, Ile, and Val and may be involved in their transme
 A:Description: Liv-binding protein
 C:Superfamily: Liv-binding protein
 F:1-42/Domain: (or 20-42) signal sequence #status predicted <SIG>
 F:43-386/Product: leucine/isoleucine/valine-binding protein #status experimental <WAT>
 F:95-120/Disulfide bonds: #status experimental

Query Match 87.1%; Score 27; DB 1; Length 386;
 Best Local Similarity 66.7%; Pred. No. 99;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 MNVGR 6
 11:11:
 Db 20 MNIKGK 25

Search completed: December 14, 2000, 10:12:56
 Job time: 149 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 14, 2000, 10:11:15 : Search time 15.82 Seconds
(without alignments)
12.117 Million cell updates/sec

Title: US-09-051-034a-11

Perfect score: 31
Sequence: 1 MNVKGK 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters: 87993

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	100.0	371	1	GATR_PIG
2	31	100.0	378	1	ASPG_CAEEL
3	28	90.3	368	1	GATR_BOVIN
4	28	90.3	394	1	GATR_MOUSE
5	28	90.3	732	1	IUTA_ECOLI
6	27	87.1	322	1	Y378_METUA
7	27	87.1	505	1	Y76J_CAEEL
8	27	87.1	559	1	PP71_HCMVA
9	27	87.1	559	1	VNUC_INBLE
10	27	87.1	2481	1	UN52_CAEEL
11	26	83.9	149	1	YOBN_BACSU
12	26	83.9	184	1	YPE3_RHORU
13	26	83.9	259	1	VP33_NPVAC
14	26	83.9	271	1	T2S1_LALIC
15	26	83.9	282	1	VP33_NPVOP
16	26	83.9	375	1	Y785_METUA
17	26	83.9	560	1	VNUC_INBAA
18	26	83.9	560	1	VNUC_INBAC
19	26	83.9	560	1	VNUC_INBAD
20	26	83.9	560	1	VNUC_INBP9
21	26	83.9	560	1	VNUC_INBSI
22	26	83.9	640	1	TRAG_AGR6
23	26	83.9	658	1	TRAG_AGR75
24	26	83.9	1451	1	MYM1_HUMAN
25	26	80.6	208	1	FGF6_HUMAN
26	26	80.6	208	1	FGF6_MOUSE
27	26	80.6	220	1	Y008_METUA
28	26	80.6	221	1	Y008_METUA
29	26	80.6	227	1	PRL_HUMAN
30	26	80.6	227	1	PRL_MOUSE
31	26	80.6	256	1	FEF3_BRARE
32	26	80.6	284	1	NODO_RHILV
33	26	80.6	302	1	YHCT_BACSU

34	25	80.6	367	1	LIV1_CITFR
35	25	80.6	378	1	SUCC_THEFL
36	25	80.6	391	1	PEXG_YARLI
37	25	80.6	416	1	PGK_HORSE
38	25	80.6	417	1	ASPG_METUA
39	25	80.6	468	1	F263_HUMAN
40	25	80.6	468	1	F263_RAT
41	25	80.6	469	1	F26L_CHICK
42	25	80.6	472	1	SACB_BACAM
43	25	80.6	493	1	ACCD_EPIVI
44	25	80.6	524	1	SRP_SOYBN
45	25	80.6	566	1	CO9_FUGRU

ALIGNMENTS

RESULT	ID	GATR_PIG	STANDARD	PRT	371 AA.
AC	P50127				
DT	01-OCT-1996 (Rel. 34, Last sequence update)				
DT	01-OCT-1996 (Rel. 34, Last annotation update)				
DE	N-ACETYLGLUCOSAMINIDE ALPHA-1,3-GALACTOSYLTRANSFERASE (EC 2.4.1.151)				
DE	(GALACTOSYLTRANSFERASE) (UDP-GALACTOSE:BETA-D-GALACTOSYL-1,4-N-ACETYL-D-GALACTOSAMINIDE ALPHA-1,3-GALACTOSYLTRANSFERASE).				
GN	GCTAL				
OS	Sus scrofa (Pig).				
OC	Eukaryota: Metazoa: Chordata: Vertebrata: Euteleostomi:				
OC	Mammalia: Eutheria: Cetartiodactyla: Suina: Suidae: Sus.				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN-YORKSHIRE: TISSUE-SPLEEN:				
RX	MEDLINE: 95104914.				
RA	Stratman K.M., Gu F., Preece A.F., Gustavsson I., Andersson L., Gustafsson K.:				
RT	"CDNA sequence and chromosome localization of pig alpha 1.3 galactosyltransferase."				
RL	Immunogenetics 41:101-105(1995).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE-SPLEEN, AND LIVER:				
RA	Sandrin M.S., Dabkowski P.L., Henning M.M., Moutouris E., McKenzie I.F.C.:				
RL	Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.				
CC	-1- FUNCTION: TRANSFER OF GALACTOSE FROM UDP-GALACTOSE TO AN ACCEPTOR MOLECULE (R).				
CC	-1- CATALYTIC ACTIVITY: UDP-GALACTOSE + BETA-D-GALACTOSYL-1,4-N-ACETYL-D-GALACTOSAMINYL-R = UDP + ALPHA-D-GALACTOSYL-1,3-BETA-D-GALACTOSYL-1,4-N-ACETYL-D-GALACTOSAMINYL-R.				
CC	-1- COFACTOR: MANGANESE.				
CC	-1- PATHWAY: GLYCOSYLATION.				
CC	-1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND				
CC	-1- FORM IN TRANS CISTERNAE OF GOLGI.				
CC	-1- SIMILARITY: STRONG, TO BLOOD GROUP AB TRANSFERASE.				
CC	-1- SIMILARITY: STRUCTURAL SIMILARITY WITH THE OTHER MAMMALIAN GLYCOSYLTRANSFERASES.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL: L36152; AAA73558.1; -				
DR	EMBL: L36535; AAA58775.1; -				
KW	Transferase: Glycosyltransferase: Glycoprotein: Transmembrane: Signal-anchor: Golgi stack: Alternative splicing: Manganese. CYTOPLASMIC (POTENTIAL).				
FT	DOMAIN 1 22 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).				
FT	TRANSMEM 7 22				

FT DOMAIN 23 371 LUMENAL, CATALYTIC (POTENTIAL).
 FT CARBOHYD 59 59 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 296 296 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 27 39 SPEGLFWINOSK -> R (IN SHORT ISOFORM).
 FT CONFLICT 227 227 M -> I (IN REF. 2).
 SO SEQUENCE 371 AA: 43764 MW: CFC715EB8D993D4 CRC64:

Query Match 100.0%; Score 31; DB 1; Length 371;
 Best Local Similarity 100.0%; Pred. No. 2,6;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVGR 6
 111111
 DB 1 MNVGR 6

RESULT 2
 ASPG_CAEEL STANDARD; PRT; 378 AA.
 ID 021697;
 AC 021697;
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE PUTATIVE N4-(BETA-N-ACETYLGLUCOSAMINYL)-L-ASPARAGINASE PRECURSOR
 DE (EC 3.5.1.26) (GLYCOSYLASPARAGINASE) (ASPARTYLGLUCOSAMINIDASE) (N4-(N-
 DE ACETYL-BETA-GLUCOSAMINYL)-L-ASPARAGINE AMIDASE) (N4).
 GN R04B3.2.
 OS Ctenorhynchus elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderidae; Caenorhabditis.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Paulley A., Gattung S.;
 RL Submitted (Mar-1986) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: CLEAVES THE GLCNAC-ASN BOND WHICH JOINS OLIGOSACCHARIDES
 CC -1- TO THE PEPTIDE OF ASPARAGINE-LINKED GLYCOPROTEINS (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: N4-(BETA-N-ACETYL-D-GLUCOSAMINYL)-L-ASPARAGINE
 CC + H(2O) -> N-ACETYL-BETA-GLUCOSAMINYLAMINE + L-ASPARATE.
 CC -1- SUBCELLULAR LOCATION: LYSOSOMAL (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE ASPARAGINASE 2 FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U50198; AAA91260.1; .
 DR HSSP: P20933; IAPZ.
 DR W04BEP; R04B3.2; CE07412.
 DR INTERPRO: IPR000246; .
 DR Pfam: PF0112; Asparaginase-2; 1.
 KW Hypothetical protein; Glycoprotein; Hydrolyase; Lysosome; Signal.
 FT SIGNAL 1 19
 FT CHAIN 20 226 GLYCOSYLASPARAGINASE, ALPHA CHAIN (BY
 FT CHAIN 227 378 SIMILARITY).
 FT CHAIN 227 378 GLYCOSYLASPARAGINASE, BETA CHAIN (BY
 FT CHAIN 227 378 SIMILARITY).
 FT DISULFID 62 67 BY SIMILARITY.
 FT DISULFID 176 192 BY SIMILARITY.
 FT DISULFID 337 361 BY SIMILARITY.
 FT ACET-SITE 227 227 BY SIMILARITY.
 FT CARBOHYD 15 15 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 133 133 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 211 211 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 351 351 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 378 AA: 41012 MW: C9694965167DCE6A CRC64:

Query Match 100.0%; Score 31; DB 1; Length 378;
 Best Local Similarity 100.0%; Pred. No. 2,6;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVGR 6
 111111
 DB 327 MNVGR 332

RESULT 3
 GATR_BOVIN STANDARD; PRT; 368 AA.
 ID GATR_BOVIN
 AC P14769;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE N-ACETYLGLUCOSAMINIDE ALPHA-1,3-GALACTOSYLTRANSFERASE (EC 2.4.1.151)
 DE (GALACTOSYLTRANSFERASE) (UDP-GALACTOSE:BETA-D-GALACTOSYL-1,4-N-ACETYL-
 DE D-GLUCOSAMINIDE ALPHA-1,3-GALACTOSYLTRANSFERASE).
 GN GCTA1.
 OS Bos taurus (Bovine).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 OC Bovidae; Bovinae; Bos.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC MEDLINE: 89340543.
 RA Joziasse D.H., Shaper J.H., van den Eljnden D.H., van Tunen A.J.,
 RA Shaper N.L.;
 RT "Bovine alpha 1->3-galactosyltransferase: isolation and
 RT characterization of a cDNA clone. Identification of homologous
 RT sequences in human genomic DNA."
 RJ J. Biol. Chem. 264:14290-14297(1989).
 CC -1- FUNCTION: TRANSFER OF GALACTOSE FROM UDP-GALACTOSE TO AN
 CC ACCEPTOR MOLECULE (R).
 CC -1- CATALYTIC ACTIVITY: UDP-GALACTOSE + BETA-D-GALACTOSYL-1,4-
 CC N-ACETYL-D-GLUCOSAMINYL-R = UDP + ALPHA-D-GALACTOSYL-1,3-
 CC BETA-D-GALACTOSYL-1,4-N-ACETYL-D-GLUCOSAMINYL-R.
 CC -1- COFACTOR: MANGANESE.
 CC -1- PATHWAY: GLYCOSYLATION.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
 CC FORM IN TRANS CISTERNAE OF GOLGI.
 CC -1- DISEASE: AUTOIMMUNE DISEASE (ANTIBODIES AGAINST ENZYMAIC
 CC PRODUCT).
 CC -1- SIMILARITY: STRONG, TO BLOOD GROUP AB TRANSFERASE.
 CC -1- SIMILARITY: STRUCTURAL SIMILARITY WITH THE OTHER MAMMALIAN
 CC GLYCOSYLTRANSFERASES.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: J04989; AAA30558.1; .
 DR PIR: A44785.
 KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;
 KW Signal-anchor; Golgi stack; Manganese.
 FT DOMAIN 1 6 CYTOPLASMIC (POTENTIAL).
 FT TRANSSEM 7 22 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).
 FT DOMAIN 23 368 LUMENAL, CATALYTIC (POTENTIAL).
 FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SO SEQUENCE 368 AA: 43246 MW: 5BC50D6737BDC33 CRC64;

Query Match 90.3%; Score 28; DB 1; Length 368;
 Best Local Similarity 83.3%; Pred. No. 15;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVGR 6
 111111

DB 1 MANYGK 6

RESULT 4

ID GATR_MOUSE STANDARD: PRT: 394 AA.

AC P23336;

DT 01-NOV-1991 (Rel. 20, Created)

DT 01-NOV-1991 (Rel. 20, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE N-ACETYLGLUCOSAMINIDE ALPHA-1,3-GALACTOSYLTRANSFERASE (EC 2.4.1.151)

DE (GALACTOSYLTRANSFERASE) (UDP-GALACTOSE:BETA-D-GALACTOSYL-1,4-N-ACETYL-D-GLUCOSAMINIDE ALPHA-1,3-GALACTOSYLTRANSFERASE).

GN GGT1 OR GGT1-1.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

NC [1]

RP SEQUENCE FROM N.A.

RX MEDLINE: 90046769.

RA Larsen R.D., Rajan V.P., Ruff M.M., Kukowska-Latallo J., Cummings R.D., Lowe J.B.,

RT "Isolation of a cDNA encoding a murine UDPgalactose:beta-D-galactosyl-1,4-N-acetyl-D-glucosaminide alpha-1,3-galactosyltransferase: expression cloning by gene transfer."

RT Proc. Natl. Acad. Sci. U.S.A. 86:8227-8231(1989).

RN [2]

RP SEQUENCE FROM N.A.

RX MEDLINE: 92184813.

RA Joziasse D.H., Shaper N.L., Kim D., van den Eijnden D.H., Shaper J.H.,

RT "Murine alpha 1,3-galactosyltransferase. A single gene locus specifies four isoforms of the enzyme by alternative splicing."

RT J. Biol. Chem. 267:5534-5541(1992).

CC -1- FUNCTION: TRANSFER OF GALACTOSE FROM UDP-GALACTOSE TO AN ACCEPTOR MOLECULE (R).

CC -1- CATALYTIC ACTIVITY: UDP-GALACTOSE + BETA-D-GALACTOSYL-1,4-N-ACETYL-D-GLUCOSAMINYL-R -> UDP + ALPHA-D-GALACTOSYL-1,3-BETA-D-GALACTOSYL-1,4-N-ACETYL-D-GLUCOSAMINYL-R.

CC -1- COFACTOR: MANGANESE.

CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND

CC -1- FORM IN TRANS CISTERNAE OF GOLGI.

CC -1- SIMILARITY: STRONG, TO BLOOD GROUP AB TRANSFERASE.

CC -1- SIMILARITY: STRUCTURAL SIMILARITY WITH THE OTHER MAMMALIAN GLYCOSYLTRANSFERASES.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC -----

CC EMBL: M26925; AAA37657.1; -

CC EMBL: M65153; AAA37711.1; -

DR PIR: A34417; A34417.

DR MGI: 95704; GGT1.

KW Transferase; Glycosyltransferase; Glycoprotein; Transmembrane; Signal-anchor; Golgi stack; Alternative splicing; Manganese.

KM

FT DOMAIN 1 41 CYTOPLASMIC (POTENTIAL).

FT TRANSMEM 42 60 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN).

FT DOMAIN 61 394 LUMENAL, CATALYTIC (POTENTIAL).

FT CARBOHYD 83 83 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT CARBOHYD 319 319 N-LINKED (GLCNAC. . .) (POTENTIAL).

FT VARSPLIC 1 35 MISSING (IN SHORTER ISOFORM).

FT VARSPLIC 62 62 R -> SPDSFLMIVHTK (IN SHORTER ISOFORM).

SO SEQUENCE 394 AA: 46475 MW: 7766831640DIBBF7 CRC64;

Query Match 90.3%; Score 28; DB 1; Length 394;

Best Local Similarity 83.3%; Pred. No. 16;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MANYGK 6

DB 36 MANYGK 41

RESULT 5

ID IUTR_ECOLI STANDARD: PRT: 732 AA.

AC P14342;

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-NOV-1995 (Rel. 32, Last sequence update)

DT 15-JUL-1999 (Rel. 38, Last annotation update)

DE FERRIC AEROBACTIN RECEPTOR PRECURSOR (CLOACIN RECEPTOR).

GN IUTR.

OS Escherichia coli.

OC Bacteri; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

NC [1]

RP SEQUENCE FROM N.A.

RX STRAIN-F344;

RA Krone W.J.A., Stegheuis F., Koningsstein G., van Doorn C., Roosenendaal B., de Graaf F.K., Oudega B.;

RT "Characterization of the pCOLV-K30 encoded cloacin Df13/aerobactin outer membrane receptor protein of Escherichia coli: isolation and purification of the protein and analysis of its nucleotide sequence and primary structure."

RT FEWS Microbiol. Lett. 26:153-161(1985).

RN [2]

RP REVISIONS.

RA Oudega B.;

RT Submitted (FEB-1992) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: RECEPTOR FOR CLOACIN DF13/AEROBACTIN.

CC -1- SUBCELLULAR LOCATION: OUTER MEMBRANE.

CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC -----

CC EMBL: X05874; CAA29297.1; -

CC EMBL: X05874; CAA29298.1; ALT_INIT.

DR PIR: S01042; S01042.

DR INTERPRO: IPR000531; -

DR PFAM: PF00593; TonB_boxC.1.

DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; 1.

DR PROSITE: PS01156; TONB_DEPENDENT_REC_2; 1.

KW Outer membrane; Iron transport; Transport; TonB box; Signal; Plasmid; Receptor.

KM

FT SIGNAL 1 25

FT CHAIN 26 732 FERRIC AEROBACTIN RECEPTOR.

FT SIMILAR 31 38 TONB BOX.

FT SIMILAR 715 732 TONB C-TERMINAL BOX.

SO SEQUENCE 732 AA: 81014 MW: 0C23879C0B7AE2B CRC64;

Query Match 90.3%; Score 28; DB 1; Length 732;

Best Local Similarity 83.3%; Pred. No. 30;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MANYGK 6

DB 92 MANYGK 97

RESULT 6

Y378_METUA

```

ID V378.META STANDARD: PRT: 322 AA.
AC 057823;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE HYPOTHETICAL PROTEIN M0378.
GN M0378.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE; 96337999.
RA Blitt C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kesteven G.A., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Seelt J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii".
RL Science 273:1058-1073(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: U67490; AAB98367.1; -.
DR TRIM: M10378; -.
DR INTERPRO: IPR002729; -.
DR PIR: P01867; DUF48; 1.
DR Hypothetical protein.
KW SEQUENCE 322 AA; 38147 MW; 6E1833962A77EA3D CRC64;
SQ
Query Match 87.1%; Score 27; DB 1; Length 322;
Best Local Similarity 83.3%; Pred. No. 23;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 1 MNVGR 6
Db 146 MNVEGR 151

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
DR EMBL: Z81505; CAB04121.1; ALT-SEQ.
DR MORPEP: F16A11.2; CE09424.
DR INTERPRO: IPR001233; -.
DR PIR: P01139; UPF0027; 1.
DR PROSITE: PS01288; UPF0027; 1.
KW Hypothetical protein.
SQ SEQUENCE 505 AA; 55230 MW; D528F702E2586909 CRC64;
Query Match 87.1%; Score 27; DB 1; Length 505;
Best Local Similarity 83.3%; Pred. No. 36;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 1 MNVGR 6
Db 30 MNVEGR 35

```



```

FT DOMAIN 1887 1976 IG-LIKE C2-TYPE DOMAIN 11.
FT DOMAIN 1977 2070 IG-LIKE C2-TYPE DOMAIN 12.
FT DOMAIN 2078 2162 IG-LIKE C2-TYPE DOMAIN 13.
FT DOMAIN 2174 2260 IG-LIKE C2-TYPE DOMAIN 14.
FT DOMAIN 2267 2347 IG-LIKE C2-TYPE DOMAIN 15.
FT DOMAIN 2354 2436 IG-LIKE C2-TYPE DOMAIN 16.
FT DMSULFID 66 114 BY SIMILARITY.
FT DMSULFID 149 161 BY SIMILARITY.
FT DMSULFID 156 174 BY SIMILARITY.
FT DMSULFID 168 183 BY SIMILARITY.
FT DMSULFID 190 202 BY SIMILARITY.
FT DMSULFID 197 215 BY SIMILARITY.
FT DMSULFID 209 224 BY SIMILARITY.
FT DMSULFID 233 246 BY SIMILARITY.
FT DMSULFID 240 259 BY SIMILARITY.
FT DMSULFID 253 268 BY SIMILARITY.
FT DMSULFID 954 963 BY SIMILARITY.
FT DMSULFID 956 970 BY SIMILARITY.
FT DMSULFID 973 982 BY SIMILARITY.
FT DMSULFID 985 1001 BY SIMILARITY.
FT DMSULFID 1010 1020 BY SIMILARITY.
FT DMSULFID 1012 1026 BY SIMILARITY.
FT DMSULFID 1029 1038 BY SIMILARITY.
FT DMSULFID 1041 1057 BY SIMILARITY.
FT DMSULFID 1060 1068 BY SIMILARITY.
FT DMSULFID 1062 1078 BY SIMILARITY.
FT DMSULFID 1081 1090 BY SIMILARITY.
FT DMSULFID 1093 1108 BY SIMILARITY.
FT DMSULFID 1151 1199 BY SIMILARITY.
FT DMSULFID 1337 1383 BY SIMILARITY.
FT DMSULFID 1434 1480 BY SIMILARITY.
FT DMSULFID 1526 1572 BY SIMILARITY.
FT DMSULFID 1617 1662 BY SIMILARITY.
FT DMSULFID 1718 1766 BY SIMILARITY.
FT DMSULFID 1813 1860 BY SIMILARITY.
FT DMSULFID 1906 1953 BY SIMILARITY.
FT DMSULFID 1997 2052 BY SIMILARITY.
FT DMSULFID 2098 2146 BY SIMILARITY.
FT DMSULFID 2194 2241 BY SIMILARITY.
FT DMSULFID 2283 2328 BY SIMILARITY.
FT DMSULFID 2373 2419 BY SIMILARITY.
FT CARBOHYD 1421 1421 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 2481 AA; 270812 MW; E47FBE2B310F5031 CRC64;

```

Query Match 87.1%; Score 27; DB 1; Length 2481;
 Best Local Similarity 83.3%; Pred. No. 1.9e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVGR 6
 Db 347 INVGR 352

```

RESULT 11
YOBN_BACSU STANDARD; PRT; 149 AA.
ID YOBN_BACSU
AC PAS930;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE HYPOHETICAL 17.1 KDA PROTEIN IN SPOIIC-CWLA INTERGENIC REGION.
GN YOBN.
OS Bacillus subtilis.
OC Bacillus/Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RX MEDLINE; 95219086;
RA Takemaru K.-I., Mizuno M., Sato T., Takeuchi M., Kobayashi Y.;
RT "Complete nucleotide sequence of a skin element excised by DNA
rearrangement during sporulation in Bacillus subtilis."

```

```

RL Microbiology 141:323-327(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RA Kobayashi Y., Mizuno M., Masuda S., Takemaru K., Hosono S., Sato T.,
RA Takeuchi M.;
RL Submitted (MAY-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP IDENTIFICATION.
RX MEDLINE; 96084975.
RA Medigue C., Moszer I., Viari A., Danchin A.;
RT "Analysis of a Bacillus subtilis genome fragment using a co-operative
computer system prototype."
RL Gene 165:GC37-GC51(1995).
CC -1- SIMILARITY: STRONG, TO B.SUBTILIS XKDN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; D32216; BAA06946.1; -
CC DR
CC EMBL; D84432; BAA12410.1; -
CC DR
CC EMBL; Z99117; CAB14545.1; -
CC DR
CC SUBTILIS; Bg11285; YOBN.
CC KW
CC HYPOTHETICAL protein.
SQ SEQUENCE 149 AA; 17137 MW; FB413CE563C534AF CRC64;

```

Query Match 83.9%; Score 26; DB 1; Length 149;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 NVKGR 6
 Db 66 NVKGR 70

```

RESULT 12
TYPE3_RHORU STANDARD; PRT; 184 AA.
ID TYPE3_RHORU
AC P23140;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-NOV-1991 (Rel. 20, Last annotation update)
DE HYPOHETICAL PROTEIN IN PETC 3 REGION (FRAGMENT).
OS Rhodospirillum rubrum.
OC Bacteria; Proteobacteria; alpha subdivision; Rhodospirillaceae;
OC Rhodospirillum.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FRI;
RX MEDLINE; 91094774.
RA Majewski C., Trebst A.;
RT "The pet genes of Rhodospirillum rubrum: cloning and sequencing of
the genes for the cytochrome bcl-complex."
RL Mol. Gen. Genet. 224:373-382(1990).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC -----
CC EMBL; X55387; CAA39062.1; -
CC DR
CC PIR; S12260; S12260.
CC KW
CC HYPOTHETICAL protein.
FT NON_TER 184

```


Query Match 83.9%; Score 26; DB 1; Length 282;
 Best Local Similarity 100.0%; Pred. No. 35;
 Matches 5; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 MNVG 5
 Db 163 MNVG 167

Search Completed: December 14, 2000, 10:14:57
 Job time: 222 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 14, 2000, 10:10:26 ; Search time 40.16 seconds
(without alignments)
13.950 Million cell updates/sec

Title: US-09-051-034A-11
Perfect score: 31
Sequence: 1 MNVGR 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 297973 seqs, 93374136 residues
Total number of hits satisfying chosen parameters: 297973

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries.

Database : SPTREMBL_14:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mmc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_protent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	30	96.8	454	3	012518	012518 saccharomyc
2	28	90.3	308	2	006809	006809 saccharomyc
3	28	90.3	376	6	028855	028855 primates. n
4	28	90.3	473	1	090286	090286 pyrococcus
5	27	87.1	107	8	P92538	P92538 arabidopsis
6	27	87.1	218	10	Q9ST07	Q9ST07 gnetum parv
7	27	87.1	287	1	059586	059586 pyrococcus
8	27	87.1	303	13	09PT19	09PT19 neotenus ma
9	27	87.1	334	1	027156	027156 methanobact
10	27	87.1	395	2	09S513	09S513 loofah wlc
11	27	87.1	420	1	028025	028025 archaeoglob
12	27	87.1	686	5	P92006	P92006 caenorhabd1
13	27	87.1	1020	10	09SG10	09SG10 arabidopsi
14	27	87.1	1160	5	018261	018261 caenorhabd1
15	27	87.1	2295	5	09XTD2	09XTD2 caenorhabd1
16	27	87.1	2482	5	018263	018263 caenorhabd1
17	27	87.1	3375	5	09XT15	09XT15 caenorhabd1
18	26	83.9	51	12	039488	039488 bovine eph
19	26	83.9	119	9	080057	080057 bacterioph

20	26	83.9	130	10	09S140	09S140 arabidopsis
21	26	83.9	142	5	09V4K6	09V4K6 drosophila
22	26	83.9	177	10	038679	038679 allium asca
23	26	83.9	235	11	063068	063068 ratius norv
24	26	83.9	254	2	067126	067126 aquifex aeo
25	26	83.9	259	12	092452	092452 bombyx mori
26	26	83.9	278	2	09RRE3	09RRE3 delinococcus
27	26	83.9	343	10	09ZOM3	09ZOM3 arabidopsis
28	26	83.9	364	2	067350	067350 aquiflex aeo
29	26	83.9	364	2	09R701	09R701 agrobacteri
30	26	83.9	401	10	09XEF3	09XEF3 arabidopsis
31	26	83.9	403	12	041569	041569 human immun
32	26	83.9	405	5	018190	018190 caenorhabd1
33	26	83.9	410	10	09S125	09S125 arabidopsis
34	26	83.9	444	10	049356	049356 arabidopsis
35	26	83.9	466	2	P77286	P77286 escherichia
36	26	83.9	486	2	068384	068384 pseudomonas
37	26	83.9	506	10	064879	064879 arabidopsis
38	26	83.9	512	10	09S879	09S879 arabidopsis
39	26	83.9	560	12	082661	082661 influenza b
40	26	83.9	560	12	082662	082662 influenza b
41	26	83.9	560	12	09WD71	09WD71 influenza b
42	26	83.9	560	12	09WD70	09WD70 influenza b
43	26	83.9	560	12	09WD69	09WD69 influenza b
44	26	83.9	560	12	09WD68	09WD68 influenza b
45	26	83.9	560	12	09WD67	09WD67 influenza b

ALIGNMENTS

RESULT 1
ID 012518 PRELIMINARY: PRT: 454 AA.
AC 012518:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE ORF_YDL161W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetales; Saccharomycetes.
RN [1]
RP SEQUENCE FROM N.A.
RA Pohl T.M.:
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA MIPS:
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-ALPHAS28C:
RA Pohl T.M.:
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL: 274210; GAA98736.1; -
DR EMBL: 267750; GAA91585.1; -
DR INTERPRO: IPR001026; -
DR PFM: PF01417; ENTH: 1.
KW Hypothetical protein.
SQ SEQUENCE 454 AA: 52352 MW: 37447FBA6449256 CRC64;

Query Match 96.8% Score 30: DB 3: Length 454;
Best Local Similarity 83.3% Pred. NO. 35;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNVGR 6
DB 145 MNVGR 150

RESULT 2

006809 ID 006809 PRELIMINARY; PRT; 308 AA.
AC 006809;
DT 01-JUL-1997 (TREMREL. 04, Created)
DT 01-JUL-1997 (TREMREL. 04, Last sequence update)
DT 01-JUN-2000 (TREMREL. 14, Last annotation update)
DE GENE.
GN GENE OR RV1451 OR MTCY493.03C.
OS Mycobacterium tuberculosis.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
RN 01
RP SEQUENCE FROM N.A.
RC SPRAIN-H37RV;
RX MEDLINE: 98295987.
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hendby T., Jags K., Kiroh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rafter S., Seeger K., Skelton S., Squares S., Squires R., Sulston J.E.,
RA Taylor K., Whitehead S., Barrell B.G.;
RT Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence."
RL Nature 393:537-544(1998).
DR EMBL: Z95844; CAB09270.1; -.
DR TUBERCULIST: RV1451; -.
DR INTERPRO: IPR000537; -.
DR PIR: P01040; UblA; 1
SQ SEQUENCE 308 AA; 33538 MW; 9879539CABD2B47E CRC64;

Query Match 90.3%; Score 28; DB 2; Length 308;
Best Local Similarity 83.3%; Pred. No. 70;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVGR 6
Db 1 MNVGR 6

RESULT 3
ID 028855 PRELIMINARY; PRT; 376 AA.
AC 028855;
DT 01-FEB-1997 (TREMREL. 02, Created)
DT 01-FEB-1997 (TREMREL. 02, Last sequence update)
DT 01-NOV-1998 (TREMREL. 08, Last annotation update)
DE N-ACETYLGLUCOSAMINIDE ALPHA-1,3-GALACTOSYLTRANSFERASE (EC 2.4.1.151)
DE (GALACTOSYLTRANSFERASE)
DE (UDP-GALACTOSE:BETA-D-GALACTOSYL-1,4-N-ACETYL-D-GLUCOSAMINIDE
DE ALPHA-1,3-GALACTOSYLTRANSFERASE).
OS Primates.
OC Eumetazoa; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria.
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE: 94331837.
RA Hendon T.R., Macher B.A., Anaraki F., Gallili U.;
RT Defining the minimal size of catalytically active primate alpha 1,3
RT galactosyltransferase: structure-function studies on the recombinant
RT truncated enzyme."
RL Glycobiology 4:193-201(1994).
- FUNCTION: TRANSFER OF GALACTOSE FROM UDP-GALACTOSE TO AN ACCEPTOR
- MOLECULE (R).
- CATALYTIC ACTIVITY: UDP-GALACTOSE + BETA-D-GALACTOSYL-1,4-N-
ACETYL-D-GLUCOSAMINYL-R = UDP + ALPHA-D-GALACTOSYL-1,3-BETA-D-
GALACTOSYL-1,4-N-ACETYL-D-GLUCOSAMINYL-R.
- COFACTOR: MANGANESE.
- PATHWAY: GLYCOSYLATION.
- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
- FORM IN TRANS CISTERNAE OF GOLGI.

CC -1- SIMILARITY: STRONG, TO BLOOD GROUP AB TRANSFERASE.
CC -1- SIMILARITY: STRUCTURAL SIMILARITY WITH THE OTHER MAMMALIAN
CC GLYCOSYLTRANSFERASES.
DR EMBL: S71333; AAB31587.1; -.
KW Transferase; Glycosyltransferase; Signal-anchor; Transmembrane;
KW Manganese; Glycoprotein.
FT DOMAIN 1 6
FT TRANSMEM 1 22
FT SIGNAL-ANCHOR
FT (TYPE-II MEMBRANE PROTEIN).
FT LUMENAL, CATALYTIC (POTENTIAL).
FT POTENTIAL.
FT CARBOHYD 301 301
FT MUTAGEN 1 46
FT MUTAGEN 1 89
FT MUTAGEN 1 90
FT MUTAGEN 1 91
FT MUTAGEN 1 92
FT MUTAGEN 1 93
FT MUTAGEN 1 96
FT MUTAGEN 1 101
FT MUTAGEN 318 318
FT MUTAGEN 374 376
FT MUTAGEN 375 375
FT MUTAGEN 375 375
FT MUTAGEN 376 376
SQ SEQUENCE 376 AA; 44471 MW; 13C490C0BE2FA2DF CRC64;

Query Match 90.3%; Score 28; DB 6; Length 376;
Best Local Similarity 83.3%; Pred. No. 86;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVGR 6
Db 1 MNVGR 6

RESULT 4
ID 0902H6 PRELIMINARY; PRT; 473 AA.
AC 0902H6;
DT 01-MAY-2000 (TREMREL. 13, Created)
DT 01-MAY-2000 (TREMREL. 13, Last sequence update)
DT 01-JUN-2000 (TREMREL. 14, Last annotation update)
DE HYPOTHETICAL 54.9 KDA PROTEIN.
GN PAB0780.
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-ORSAY;
RA Hellig R.;
RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome
RT structure and evolution."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: A0248286; CAB50082.1; -.
KW Hypothetical protein.
SQ SEQUENCE 473 AA; 54913 MW; 40D8E832AE9BC24A CRC64;

Query Match 90.3%; Score 28; DB 1; Length 473;
Best Local Similarity 83.3%; Pred. No. 11e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVGR 6
Db 331 MNVGR 336

RESULT 5
ID P92538 PRELIMINARY; PRT; 107 AA.
AC P92538;
DT 01-MAY-1997 (TREMREL. 03, Created)

```

DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)
DE 01-MAY-1997 (TREMBLrel. 03, Last annotation update)
DE 08F107F.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Mitochondrion.
OC Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;
OC Brassicaceae; Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA Unseid M., Martenfeld J.R., Brandt P., Brennick A.;
RL Nat. Genet. 0:0-0(0).
DR EMBL: Y08502; CAA69790.1; -.
KM Mitochondrion.
SQ SEQUENCE 107 AA; 12130 MW; 673FA5BB44E855E CRC64;

Query Match      87.1%; Score 27; DB 8; Length 107;
Best Local Similarity 83.3%; Pred. No. 42;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNVGR 6
DB 10 MNVGR 15

RESULT 6
ID 09ST07 PRELIMINARY; PRT; 218 AA.
AC 09ST07;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE GPMADSL PROTEIN.
GN GPMADSL.
OS Gnetum parvifolium.
OC Eukaryota: Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Gnetales; Gnetales; Gnetales; Gnetales; Gnetales; Gnetales;
RN [1]
RP SEQUENCE FROM N.A.
RA Shindo S., Ito M., Ueda K., Kato M., Hasebe M.;
RT "Characterization of MADS genes in the gymnosperm Gnetum parvifolium
RT and its implication on the evolution of reproductive organs in seed
RT plants."
RL Eviol. Dev. 1:1-11(1999).
DR EMBL: AB022644; BA85629.1; -.
DR INTERPRO: IPR002100; -.
DR INTERPRO: IPR002487; -.
DR PFAM: PF00319; SRF-TF; 1.
DR PFAM: PF01486; K-box; 1.
DR PRINTS: PR00404; MADSOMAIN.
DR PROSITE: PS50066; MADS_BOX_2; 1.
SQ SEQUENCE 218 AA; 24908 MW; 7449362FCA836F04 CRC64;

Query Match      87.1%; Score 27; DB 10; Length 218;
Best Local Similarity 83.3%; Pred. No. 86;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNVGR 6
DB 155 MNVGR 160

RESULT 7
ID 059586 PRELIMINARY; PRT; 287 AA.
AC 059586;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE 287AA LONG HYPOTHEICAL RIBOSE-PROSPHATE PYROPHOSPHOKINASE.
GN PH1923.

```

```

OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OT3.
RX MEDLINE: 98344137.
RA Kawarabayashi Y., Sawada M., Horikawa H., Halkava Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funehashi T., Tanaka T., Kudoh Y., Yamazaki J., Kusihide N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3."
RL DNA Res. 5:55-76(1998).
CC -1 SIMILARITY: PARTIAL WITH OTHER PHOSPHORIBOSYLTRANSFERASE ALSO
INVOLVED IN BIOSYNTHESIS OR SALVAGE OF PURINES OR PYRIMIDINES.
DR EMBL: AP000007; BA31050.1; -.
DR INTERPRO: IPR000836; -.
DR INTERPRO: IPR002375; -.
DR PFAM: PF00156; Pribosyltran; 1.
DR PROSITE: PS00103; PUR_PRR_PR_TRANSFER; 1.
KM Transferase; Glycosyltransferase.
SQ SEQUENCE 287 AA; 32094 MW; FB0870C9F1CF1235 CRC64;

Query Match      87.1%; Score 27; DB 1; Length 287;
Best Local Similarity 83.3%; Pred. No. 1;e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNVGR 6
DB 200 MNVGR 205

RESULT 8
ID 09PT19 PRELIMINARY; PRT; 303 AA.
AC 09PT19;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE ARYL HYDROCARBON RECEPTOR (FRAGMENT).
GN AHR.
OS Neoturus maculosus.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandrolidae; Salamandridae; Neoturus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-LIVER;
RA Karchner S.I., Kennedy S.W., Trudeau S., Hahn M.E.;
RT "Towards a molecular understanding of species differences in dioxin
RT sensitivity: Initial characterization of Ah receptor cDNAs in birds
RT and an amphibian."
RL Mar. Environ. Res. 0:0-0(2000).
DR EMBL: AF192500; AAF15278.1; -.
DR INTERPRO: IPR000014; -.
DR INTERPRO: IPR001092; -.
DR INTERPRO: IPR003015; -.
DR PFAM: PF00989; PAS; 1.
DR PROSITE: PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
KM Receptor.
FT NON_TER 303
FT NON_TER 303
SQ SEQUENCE 303 AA; 34172 MW; EDC8BA6110720499 CRC64;

Query Match      87.1%; Score 27; DB 13; Length 303;
Best Local Similarity 83.3%; Pred. No. 1.2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNVGR 6
DB 1 MNVGR 6

```

Db 199 MNVGR 204

RESULT 9
ID 027156 PRELIMINARY: PRT: 334 AA.

DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE CONSERVED PROTEIN.
GN MTH1084.
OS Methanobacterium thermoautotrophicum.
OC Archaea; Euryarchaeota; Methanobacteriales; Methanobacteriaceae;
CC Methanobacterium.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-DELTA H;
RX MEDLINE: 98037514.

RA Smith D.R., Doucelette-Stamm L.A., Deloughery C., Lee H.-M., Dubois J.,
RA Aldredge T., Bashirzadeh R., Blakely D., Cook R., Gilbert K.,
RA Harrison D., Hoang L., Keagle P., Lum W., Pothier B., Qiu D.,
RA Spadafora R., Vlcare R., Wang Y., Wierzbowski J., Gibson R.,
RA Jiwani N., Carnio A., Bush D., Safer H., Patwell D., Prabhakar S.,
RA McQuay S., Shimer G., Goyal A., Pietrowski S., Church G.M.,
RA Daniels C.J., Mao J.-I., Rice P., Nolling J., Reeve J.N.;
RT *Complete genome sequence of Methanobacterium thermoautotrophicum
RT delta: functional analysis and comparative genomics.";
RL J. Bacteriol. 179:7135-7155(1997).
DR EMBL: AE000879; AAB85573.1; -.
DR INTERPRO: IPR002729; -.
DR PIRAM: PF01861; DUF48; 1.
DR PRODOM: PD008695; -; 1.
SQ SEQUENCE 334 AA; 38817 MW; 25ABEF9A483285EE CRC64;

QY 1 MNVGR 6
Db 158 MNVGR 163

Query Match 87.1%; Score 27; DB 1; Length 334;
Best Local Similarity 83.3%; Pred. NO. 1.3e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

RESULT 10
ID 095513 PRELIMINARY: PRT: 395 AA.
AC 095513;
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE ELONGATION FACTOR EF-TU.
GN TUF.
OS Lodiia witches'-broom phytoplasma.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
CC Actinoplasmataceae; Phytoplasma.
RN [1]
RP SEQUENCE FROM N.A.
RA Chen Y.-H., Ho K.-C.;
RT "Nucleotide sequence of tuf gene of a phytoplasma associated with
RT Lodiia witches'-broom.";
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF086617; AAD48153.1; -.
DR INTERPRO: IPR000795; -.
DR PIRAM: PF00009; GTP_EFTU; 1.
DR PRINTS: PR00315; ELONGATNFACT.
KM Elongation factor.
SQ SEQUENCE 395 AA; 43833 MW; 979DE22A1EDF9DEB CRC64;

Query Match 87.1%; Score 27; DB 2; Length 395;
Best Local Similarity 83.3%; Pred. NO. 1.6e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNVGR 6
Db 218 MNVGR 223

RESULT 11
ID 028025 PRELIMINARY: PRT: 420 AA.

AC 028025;
DT 01-JAN-1998 (TREMblrel. 05, Created)
DT 01-JAN-1998 (TREMblrel. 05, Last sequence update)
DT 01-AUG-1998 (TREMblrel. 07, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN AF2259.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
CC Archaeoglobus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE: 98049343.

RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickley E.K., Peterson J.D.,
RA Fleischmann D.L., Kerlavage A.R., Graham D.E., Kyrpides N.C.,
RA Richardson S.F., Dougherty B.A., Lee N.H., Sutton G.G., Gill S.,
RA Kirkness E.F., Doucelette R.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.J., McNeil L.K., Badger J.H., Glodek A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Utterback T.,
RA Cotton M.D., Spriggs T., Artlich P., Kaine B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus.";
RL Nature 390:364-370(1997).
DR EMBL: AE000948; AAB88990.1; -.
DR TIGR: AF2259; -.
KM Hypothetical protein.
SQ SEQUENCE 420 AA; 48222 MW; 846D3F445EC4885 CRC64;

QY 1 MNVGR 6
Db 411 MNVGR 416

Query Match 87.1%; Score 27; DB 1; Length 420;
Best Local Similarity 83.3%; Pred. NO. 1.7e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

RESULT 12
ID P92006 PRELIMINARY: PRT: 686 AA.
AC P92006;
DT 01-MAY-1997 (TREMblrel. 03, Created)
DT 01-MAY-1997 (TREMblrel. 03, Last sequence update)
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE M04G12.3 PROTEIN.
GN M04G12.3.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
CC Rhabditidae; Petodiidae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RA Stims M.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 94150718.
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,

RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Garner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurtry A., Mortimore B., O'Callaghan M.,
 RA Parsons N., Percy C., Rikken L., Koopa A., Saunders D., Showkeen R.,
 RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierly-Mieg J., Thomas K., Vaubin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohlman P.,
 RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans.*
 RL Nature 368:32-38(1994).
 DR EMBL: 281103; CAB03210.1;
 DR HSSP: P16068; IAWN.
 DR INTERPRO: IPR001054;
 SO SEQUENCE 686 AA; 78588 MW; 65AE352DB53CBA57 CRC64;

Query Match 87.1%; Score 27; DB 5; Length 686;
 Best Local Similarity 83.3%; Pred. No. 2.8e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVGR 6
 :|||||
 DB 615 INVGR 620

RESULT 13
 O9SG10 PRELIMINARY; PRT: 1020 AA.
 AC O9SG10;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
 DE T13015.1 PROTEIN.
 GN T13015.1
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; Rosidae; euroids II; Brassicales;
 OC Brassicaceae; Arabidopsids.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B., Wu D.,
 RA Rongling C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,
 RA Bowman C.L., White O., Niernan W.C., Fraser C.M.;
 RT Arabidopsis thaliana chromosome II; BAC T13015 genomic sequence.*;
 RT Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AC010870; AAF24608.1;
 SO SEQUENCE 1020 AA; 114597 MW; 9AC3A5FAFC6EA408 CRC64;

Query Match 87.1%; Score 27; DB 10; Length 1020;
 Best Local Similarity 83.3%; Pred. No. 4.1e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVGR 6
 :|||||
 DB 932 VNVGR 937

RESULT 14
 O18261 PRELIMINARY; PRT: 1160 AA.
 AC O18261;
 DT 01-JAN-1998 (TREMBlrel. 05, Created)
 DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
 DE UNC-52 PROTEIN.
 GN UNC-52 OR ZC101.2
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Percy C.;
 RT Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- ALTERNATIVE PRODUCTS: FOUR FORMS: ISOFORM ZC101.2A (O18263),
 CC ISOFORM ZC101.2B (SHOWN HERE), ISOFORM ZC101.2C (O9XTD2) AND
 CC ISOFORM ZC101.2E (O9XTI5); MAY BE PRODUCED BY ALTERNATIVE
 CC SPLICING.
 DR EMBL: 293395; CAB07704.1;
 DR HSSP: P01130; ILDR.
 DR WORMPEP: ZC101.2B; CE15030.
 DR INTERPRO: IPR000034;
 DR INTERPRO: IPR000561;
 DR INTERPRO: IPR002049;
 DR INTERPRO: IPR002172;
 DR INTERPRO: IPR003006;
 DR PFAM: PF00052; laminin_B; 2.
 DR PFAM: PF00047; Ig; 2.
 DR PFAM: PF00053; laminin_EGF; 5.
 DR PFAM: PF00057; 1d1_recept_a; 3.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_4.
 DR PROSITE: PS01186; EGF_2; 2.
 DR PROSITE: PS01209; LDLRA_1; 3.
 DR PROSITE: PS01248; LAMININ_TYPE_EGF; 7.
 DR PROSITE: PS50068; LDLRA_2; 3.
 DR PRODOM: PD003031; -; 2.
 KW Hypothetical protein; Glycoprotein; Alternative splicing.
 SO SEQUENCE 1160 AA; 127378 MW; 571A23C709285FC6 CRC64;

Query Match 87.1%; Score 27; DB 5; Length 1160;
 Best Local Similarity 83.3%; Pred. No. 4.7e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVGR 6
 :|||||
 DB 347 INVGR 352

RESULT 15
 O9XTD2 PRELIMINARY; PRT: 2295 AA.
 AC O9XTD2;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
 DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
 DE UNC-52 PROTEIN.
 GN UNC-52 OR ZC101.2
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BRISTOL N2;
 RA Percy C., Baynes C.;
 RT Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- ALTERNATIVE PRODUCTS: FOUR FORMS: ISOFORM ZC101.2A (O18263),
 CC ISOFORM ZC101.2B (O18261), ISOFORM ZC101.2C (SHOWN HERE) AND
 CC ISOFORM ZC101.2E (O9XTI5); MAY BE PRODUCED BY ALTERNATIVE
 CC SPLICING.
 DR EMBL: 293395; CAB07707.1;
 DR EMBL: 293375; CAB07707.1; JOINED.
 DR EMBL: 293375; CAB07568.1;
 DR EMBL: 293395; CAB07568.1; JOINED.
 DR WORMPEP: ZC101.2C; CE15034.
 DR INTERPRO: IPR000034;
 DR INTERPRO: IPR000561;
 DR INTERPRO: IPR001993;
 DR INTERPRO: IPR002049;
 DR INTERPRO: IPR002172;
 DR INTERPRO: IPR003006;
 DR PFAM: PF00047; Ig; 14.
 DR PFAM: PF00052; laminin_B; 2.

DR PGM; PF00053; laminin_EGF; 5.
DR PGM; PF00057; Idl_recept_a; 3.
DR PRINTS; PR00011; EGF_LAMININ.
DR PRINTS; PR00261; LDLRECEPTOR.
DR PROSITE; PS00022; EGF_1; UNKNOWN_4.
DR PROSITE; PS00215; MITOCH_CARRIER; UNKNOWN_1.
DR PROSITE; PS01186; EGF_2; 2.
DR PROSITE; PS01209; LDLRA_1; 3.
DR PROSITE; PS01248; LAMININ_TYPE_EGF; 7.
DR PROSITE; PS50068; LDLRA_2; 3.
KW Glycoprotein; Hypothetical protein; Alternative splicing.
SQ SEQUENCE 2295 AA; 251087 MW; F6BC6067ABF86C48 CRC64;

Query Match 87.1%; Score 27; DB 5; Length 2295;
Best Local Similarity 83.3%; Pred. No. 9.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNVGR 6
:|||||
Db 347 INVGR 352

Search completed: December 14, 2000, 10:12:25
Job time: 119 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 14, 2000, 10:11:09 ; Search time 36.07 Seconds
(without alignments)
5.688 Million cell updates/sec

Title: US-09-051-034a-12

Perfect score: 31
Sequence: 1 MNVKGK 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A.GeneSeq_36:.*
1: /SIDSI/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	100.0	100	Y26039	Secreted protein n
2	31	100.0	376	R80016	Marmoset alpha-1,3-
3	31	100.0	394	R13750	GDP-Fuc:(beta-D-Ga
4	31	100.0	394	R45935	A glycosyltransfer
5	31	100.0	394	W13639	Murine alpha(1,3)-
6	28	90.3	218	W20717	H. pylori cytoplasm
7	28	90.3	342	W49689	Porcine alpha-1,3-
8	28	90.3	354	W49688	Porcine alpha-1,3-
9	28	90.3	359	R62508	Galactosyl transfer
10	28	90.3	359	R90573	Pig alpha(1,3)-gal
11	28	90.3	363	W49687	Porcine alpha-1,3-
12	28	90.3	371	R85082	Porcine alpha(1,3

13	28	90.3	371	16	R76777	Pig alpha-1,3-gala
14	28	90.3	375	19	W49686	Porcine alpha-1,3-
15	28	90.3	463	20	Y24083	Salmonella typhim
16	28	90.3	749	19	W53457	Amino acid sequenc
17	28	90.3	749	21	Y84627	Amino acid sequenc
18	27	87.1	176	21	Y74655	Neisseria meningit
19	27	87.1	185	21	Y74656	Neisseria meningit
20	27	87.1	166	21	Y74654	Neisseria gonorrhoe
21	27	87.1	429	17	R94561	Human adenyl cycl
22	27	87.1	575	20	Y20048	B. burgdorferi ant
23	27	87.1	765	19	W98803	H. pylori GHPD 130
24	27	87.1	1090	17	R94560	Rat adenyl cycl
25	27	87.1	1294	20	W30601	Human type IX aden
26	27	87.1	1305	20	W88525	Adenyl cyclase typ
27	27	87.1	1353	17	R99251	Murine adenylate c
28	26	83.9	18	21	Y86490	Human gene 59-enco
29	26	83.9	32	18	W27808	Staphylococcus aur
30	26	83.9	106	21	Y86491	Human gene 59-enco
31	26	83.9	270	19	W63115	A human pathogenes
32	26	83.9	289	21	Y67583	Maize prohibitin,
33	26	83.9	463	20	Y13392	Glucosyl IP-transf
34	26	83.9	470	17	W03997	Amino acid sequenc
35	26	83.9	470	18	W32794	Sphingomonas genus
36	26	83.9	470	20	W67750	Sphingomonas strai
37	26	83.9	647	21	Y59629	Xenopus laevis HSP
38	26	83.9	647	21	R03928	Amino acid sequenc
39	26	83.9	903	21	Y54137	P. multocida hyalu
40	26	83.9	972	20	Y06212	Rattus norvegicus
41	26	83.9	972	21	Y43099	KGF-2 deletion mut
42	26	83.9	1886	19	W54241	E. coli optimised
43	25	80.6	86	19	W52589	E. coli optimised
44	25	80.6	87	20	Y32903	E. coli optimised
45	25	80.6	92	20	Y32906	E. coli optimised

ALIGNMENTS

RESULT 1	
Y26039	
ID Y26039	Standard: Protein: 100 AA.
XX	
AC Y26039;	
XX	
DT 22-OCT-1999	(first entry)
XX	
DE Secreted protein nt746_4.	
XX	
KM Secreted protein: cytokine: cell proliferation: immune stimulation:	
KW vaccine: immune suppression: haematopoiesis: tissue growth: activin;	
KW inhibin: chemotaxis: chemokinesis: haemostasis: thrombolytic;	
KW receptor: ligand: anti-inflammatory: cadherin: tumour: gene therapy.	
XX	
OS Homo sapiens.	
XX	
PN W09928335-A1.	
XX	
PD 10-JUN-1999.	
XX	
PF 02-DEC-1998:	98WO-US25512.
XX	
PR 30-NOV-1998:	98US-0203106.
PR 04-DEC-1997:	97US-0067454.
XX	
PA (GEMV) GENETICS INST INC.	
XX	
PI Agostino MJ, Clark HF, Collins-Racie LA, Evans C;	
PI Fochtel K, Jacobs K, Lavallie ER, McCoy JM, Merberg D;	
PI Steinhilber RJ, Treacy M, Wong GG;	
DR WPI: 1999-385352/32.	
DR N-PSDB: X80674.	
XX	

PT New polynucleotides encoding secreted human proteins
 XX
 PS Glaim 34; Page 115; 124pp; English.
 XX
 CC The present sequence is a secreted protein nt746.4, encoded by known
 CC clone nt746.4 (deposited as ATCC 98600) isolated from human adult brain
 CC cDNA library. Recombinant secreted proteins can be produced by
 CC transforming host cells and culturing them under suitable conditions.
 CC The polynucleotide and protein are predicted to have biological
 CC activities which would make them suitable for treating, preventing or
 CC ameliorating medical conditions in humans and animals. Some predicted
 CC biological activities include cytokine and cell proliferation/
 CC differentiation activity, immune stimulating (e.g. as vaccines) or
 CC suppressing activity, hematopoiesis regulating activity, tissue growth
 CC activity, activin/inhibin activity, chemotactic/chemokinetic activity,
 CC hemostatic and thrombolytic activity, receptor/ligand activity, anti-
 CC inflammatory activity, cadherin/tumour invasion suppressor activity, and
 CC tumour inhibition activity. The polynucleotide encoding secreted
 CC protein can be used for gene therapy.
 XX
 SQ Sequence 100 AA;
 Query Match 100.0%; Score 31; DB 20; Length 100;
 Best Local Similarity 100.0%; Pred. No. 4.5;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 MNVKGK 6
 1
 1 mnvkgk 6
 DB
 RESULT 2
 R80016
 ID R80016 standard; Protein; 376 AA.
 XX
 AC R80016;
 XX
 DT 01-MAY-1996 (first entry)
 XX
 DE Marmoset alpha-1,3-galactosyltransferase.
 XX
 KM Marmoset; alpha-1,3-galactosyltransferase; immune response; glycoprotein;
 KM alpha-galactosyl epitope; cell membrane; vitus; phagocytosis; tumour;
 KM antigen processing; leukemia; lymphoma; myeloma; melanoma; carcinoma;
 KM sarcoma; vaccine; opsonisation; glycoprotein; antibody; anti-Gal.
 XX
 OS Gallitrix jacchus.
 XX
 PN WO9524924-A1.
 XX
 PD 21-SEP-1995.
 XX
 PE 13-MAR-1995; 95WO-US03156.
 XX
 PR 15-MAR-1994; 94US-0213200.
 XX
 PA (UWHA-) UNIV HAHNEMANN & MEDICAL COLLEGE PENNSYL.
 XX
 PI Galili U, Repik PM;
 XX
 DR WPI; 1995-336816/43.
 XX
 DR N-PSDB; T04522.
 XX
 PT Association of an alpha-galactosyl epitope with a tumour or viral
 PT antigen - is administered to anti-Gal synthesising animals to induce
 PT an immune response
 XX
 PS Disclosure; Fig 9; 85pp; English.
 XX
 CC The amino acid sequence of the marmoset alpha-1,3-galactosyltransferase.
 CC The enzyme can be used in methods of enhancing an immune response by
 CC associating the alpha-galactosyl epitope with a cell membrane or viral

CC glycoprotein. The alpha-galactosyl epitope enhances phagocytosis and
 CC subsequent processing of the antigen. The method is useful in the
 CC treatment of tumours e.g. leukemia, lymphoma, myeloma, melanoma,
 CC carcinoma and sarcoma, or for the generation of viral vaccines by
 CC opsonising a viral glycoprotein. The alpha-galactosyl epitope enhances
 CC recognition of the antigen in an animal that synthesises the naturally
 CC occurring antibody - anti-Gal.
 XX
 SQ Sequence 376 AA;
 Query Match 100.0%; Score 31; DB 16; Length 376;
 Best Local Similarity 100.0%; Pred. No. 18;
 Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 MNVKGK 6
 1
 1 mnvkgk 6
 DB
 RESULT 3
 R13750
 ID R13750 standard; Protein; 394 AA.
 XX
 AC R13750;
 XX
 DT 07-NOV-1991 (first entry)
 XX
 DE GDP-Fuc:[beta-D-Gal(1,4/1,3)]-D-GLcNAc/(Glc)alpha(1,3/1,4)
 DE -fucosyltransferase.
 DE
 XX Glycosyltransferase.
 XX
 OS Mus musculus.
 XX
 PN WO9112340-A.
 XX
 PD 22-AUG-1991.
 XX
 PE 14-FEB-1991; 91WO-US00899.
 XX
 PR 12-DEC-1990; 90US-0627621.
 PR 14-FEB-1990; 90US-0479858.
 PR 14-FEB-1990; 90US-0460133.
 XX
 PA (UNMI) UNIV OF MICHIGAN.
 XX
 PI Lowe JB;
 XX
 DR WPI; 1991-267151/36.
 DR N-PSDB; Q13331.
 XX
 PT Isolation of gene conveying post-translational characteristic -
 PT e.g. the presence of soluble or membrane bound oligo or
 PT polysaccharide or glycosyltransferase.
 XX
 PS Disclosure; Fig 2; 155pp; English.
 XX
 CC The amino acid sequence codes for a protein capable of functioning
 CC as a UDP-Gal:[beta-D-Gal(1,4)]-D-GlcNAc alpha (1,3)galacto-
 CC syltransferase. The products of this enzyme, sub-terminal alpha
 CC (1,3) and alpha(1,4) fucose residues are used in the post-
 CC translational modification of the oligosaccharides on cell surface,
 CC intracellular or secreted proteins or lipids. These can be used for
 CC the prodn. of diagnostics and therapeutics. There is a single
 CC transmembrane domain consisting of a 19 amino acid hydrophobic
 CC segment flanked by basic residues and a large (presumably
 CC catalytic) C-terminal domain that would ultimately be targeted to
 CC the lumen of the Golgi. It has two potential N-glycosylation sites
 CC indicating that as with other glycosyltransferases, it may be
 CC synthesised as a glycoprotein. It is representative of a Type II
 CC transmembrane protein. See also R13749-R13752.
 CC
 XX

```

SQ Sequence 394 AA:

Query Match 100.0%; Score 31; DB 12; Length 394;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKGK 6
DB 36 mnvkgk 41

RESULT 4
R45935
ID R45935 standard; Protein: 394 AA.
XX AC R45935;
XX DT 26-JUL-1994 (first entry)
XX DE A glycosyltransferase.
XX KW Glycosyltransferase; fucosyltransferase; GDP-Fuc: in vitro; cell;
XX KM surface; oligosaccharide.
XX OS Homo sapiens.
XX PN W09402616-A.
XX PD 03-FEB-1994.
XX PF 20-JUL-1993; 93WO-US06703.
XX PR 20-JUL-1992; 92US-0914281.
XX PA (UNMI ) UNIV MICHIGAN.
XX PI Lowe JB;
XX PI WPI: 1994-048874/06.
XX DR N-PSDB; Q56907.
XX PT DNA fragment encoding a glycosyltransferase - can be used for in
XX PR vitro reactions to modify cell surface oligosaccharide(s) e.g.
XX PS blood gp. determinants, to protect against transplant rejection
XX PS Disclosure; Fig 2; 249pp; English.
XX CC The sequence is that of a human glycosyl transferase. The enzyme
XX CC may be non glycosylated. This prevents premature loss of enzyme
XX CC activity. It can also be used in in vitro reactions to modify cell
XX CC surface oligosaccharide mols. e.g. blood group determinants.
XX CC See also R45933-9.
XX SQ Sequence 394 AA:

Query Match 100.0%; Score 31; DB 15; Length 394;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKGK 6
DB 36 mnvkgk 41

RESULT 5
W13639
ID W13639 standard; Protein: 394 AA.
XX AC W13639;
XX DT 19-JUN-1997 (first entry)

```

```

XX DE Murine alpha(1,3)-galactosyltransferase.
XX KM Alpha(1,3)-galactosyltransferase; glycosylation; oligosaccharide.
XX OS Mus sp.
XX PN W09709421-A1.
XX PD 13-MAR-1997.
XX PF 06-SEP-1996; 96WO-US13816.
XX PR 08-SEP-1995; 95US-0525058.
XX PA (UNMI ) UNIV MICHIGAN.
XX PI Legault DJ, Lowe JB;
XX PI WPI: 1997-192897/17.
XX DR N-PSDB; T61676.
XX PT New recombinant fucosyltransferase proteins - useful for modifying
XX PT cell surface oligosaccharide structures
XX PS Example 2; Page 272-274; 329pp; English.
XX CC Murine UDP-Gal:beta-D-Gal(1,4)-D-glucNAc alpha(1,3)-
XX CC galactosyltransferase (W13639) catalyses the a transglycosylation
XX CC reaction between UDP-Gal and N-acetyllactosamine and is associated
XX CC with surface-localised expression of Gal(alpha1-3)Gal linkages.
XX CC Its amino acid sequence was deduced from a cDNA clone (T61676)
XX CC obtd. by transfecting COS-1 cells with cDNA derived from mouse F9
XX CC teratocarcinoma cells, and screening the transfected cells for
XX CC surface-localised Gal(alpha1-3)Gal linkages. When expressed in
XX CC animal cell lines, the enzyme provides specific capabilities with
XX CC respect to post-translational modification of the oligosaccharides
XX CC of expressed proteins or lipids. The enzyme can also be used to
XX CC raise antibodies and to screen for inhibitor cpds.
XX SQ Sequence 394 AA:

Query Match 100.0%; Score 31; DB 18; Length 394;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKGK 6
DB 36 mnvkgk 41

RESULT 6
W20717
ID W20717 standard; Protein: 218 AA.
XX AC W20717;
XX DT 15-JUL-1997 (first entry)
XX DE H. pylori cytoplasmic protein, 05cp20518orf56.
XX KW Vaccine; prevention; treatment; infection; identification;
XX KW binding compound; bacterium; life cycle; activator; bacteria;
XX KW inhibitor; duodenal ulcer disease; chronic gastritis; diagnosis;
XX KW cytoplasmic; genome replication; transcription; recombination;
XX KW repair.
XX OS Helicobacter pylori.
XX PN W09640893-A1.
XX PD 19-DEC-1996.

```

XX 06 JUN-1996; 96MO-US09122.
 PF 01-APR-1996; 96US-0630405.
 PR 07-JUN-1995; 95US-0487032.
 XX
 PA (ASTR) ASTRA AB.
 PI Belgindh OF, Smith D, Meligaerd BL;
 XX
 DR WPI: 1997-052306/05.
 DR N-PSDB; T67970.
 XX
 PT Helicobacter pylori nucleic acid sequences and related
 PT polypeptide(s) - useful for vaccines to treat or prevent H. pylori
 PT infection, and to detect Helicobacter
 XX
 PS Claim 61: Pages 1132-1133; 1481pp; English.
 XX
 CC The present sequence is a Helicobacter pylori cytoplasmic protein
 CC involved in genome replication, transcription, recombination and
 CC repair.
 CC The protein may be used in a vaccine to prevent or treat
 CC H. pylori infection or to identify H. pylori polypeptide binding
 CC compounds, useful as potential H. pylori life cycle activators or
 CC inhibitors. The genomic sequence of H. pylori (ATCC 55679) was
 CC determined from overlapping contigs generated by mechanically
 CC shearing the bacterial DNA. The sequences were analysed for ORF of
 CC at least 180 nucleotides, and the predicted coding regions defined
 CC by computer evaluation. To identify likely H. pylori antigens for
 CC vaccine development, the amino acid sequences predicted from
 CC various ORF were analysed for significant homology to other known
 CC of exported membrane proteins. Having identified and determined
 CC the sequences of interest, particular regions can be isolated from
 CC H. pylori by PCR amplification for recombinant polypeptide
 CC production, e.g. in E. coli hosts.
 XX
 SQ Sequence 218 AA:
 QY 1 MNVKGK 6
 Db 142 LNVK GK 147
 QY :|||||
 Db 142 LNVK GK 147
 RESULT 7
 ID W49689 standard; Protein: 342 AA.
 AC W49689;
 XX
 DT 10-NOV-1998 (first entry)
 DE Porcine alpha-1,3-galactosyl transferase isoform 4.
 XX
 KM Isoform: porcine; enzyme: alpha-1,3-galactosyl transferase; galactose;
 KM sugar: N-acetyllactosamine; glycoprotein; glycolipid; antibody; pig;
 KM graft tissue rejection; organ transplantation; xenotransplant.
 XX
 OS Sus scrofa.
 XX
 PN FR2751346-A1.
 PD 23-JAN-1998.
 XX
 PF 19-JUL-1996; 96FR-0009077.
 XX
 PR 19-JUL-1996; 96FR-0009077.
 XX

PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 XX
 XX Pourcel C, Soulliou JP, Vanhove B;
 XX
 DR WPI: 1998-112876/11.
 DR N-PSDB; V49456.
 XX
 PT Transgenic non-human donors of organs for human recipients -
 PT containing DNA encoding antibodies that inhibit graft rejection
 XX
 PS Claim 4: Page 43-44; 71pp; French.
 XX
 CC This sequence represents isoform 4 of the porcine enzyme
 CC alpha-1,3-galactosyl transferase (alpha-1,3-GT). The enzyme catalyses
 CC the attachment of a galactose sugar molecule on the N-acetyllactosamine
 CC moiety found on surface glycoproteins and glycolipids. These sugar
 CC molecules are partly responsible for raising anti-graft antibodies, which
 CC lead to graft tissue rejection. The invention relates to a method of
 CC inhibiting the graft rejection mechanism by introducing the sequence
 CC encoding an antibody targeted to alpha-1,3-GT into the cells of animal,
 CC especially a pig, from whom organs may be used for xenotransplants.
 CC Neutralisation of the alpha-1,3-GT leads to tissues or organs lacking
 CC the galactose on the glycoproteins and glycolipids, thus preventing
 CC induction of the rejection response.
 XX
 SQ Sequence 342 AA:
 QY 1 MNVKGK 6
 Db 5 mNVKGR 10
 QY :|||||
 Db 5 mNVKGR 10
 RESULT 8
 ID W49688 standard; Protein: 354 AA.
 AC W49688;
 XX
 DT 10-NOV-1998 (first entry)
 DE Porcine alpha-1,3-galactosyl transferase isoform 3.
 XX
 KM Isoform: porcine; enzyme: alpha-1,3-galactosyl transferase; galactose;
 KM sugar: N-acetyllactosamine; glycoprotein; glycolipid; antibody; pig;
 KM graft tissue rejection; organ transplantation; xenotransplant.
 XX
 OS Sus scrofa.
 XX
 PN FR2751346-A1.
 PD 23-JAN-1998.
 XX
 PF 19-JUL-1996; 96FR-0009077.
 XX
 PR 19-JUL-1996; 96FR-0009077.
 XX
 PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
 XX
 PI Pourcel C, Soulliou JP, Vanhove B;
 XX
 DR WPI: 1998-112876/11.
 DR N-PSDB; V49455.
 XX
 PT Transgenic non-human donors of organs for human recipients -
 PT containing DNA encoding antibodies that inhibit graft rejection
 XX
 PS Claim 4: Page 39-41; 71pp; French.
 XX

CC This sequence represents isoform 3 of the porcine enzyme
 CC alpha-1,3-galactosyl transferase (alpha-1,3-GT). The enzyme catalyses
 CC the attachment of a galactose sugar molecule on the N-acetylglucosamine
 CC moiety found on surface glycoproteins and glycolipids. These sugar
 CC molecules are partly responsible for raising anti-graft antibodies, which
 CC lead to graft tissue rejection. The invention relates to a method of
 CC inhibiting the graft rejection mechanism by introducing the sequence
 CC encoding an antibody targeted to alpha-1,3-GT into the cells of animal,
 CC especially a pig, from whom organs may be used for xenotransplants.
 CC Neutralisation of the alpha-1,3-GT leads to tissues or organs lacking
 CC the galactose on the glycoproteins and glycolipids, thus preventing
 CC induction of the rejection response.

SO Sequence 354 AA:

Query Match 90.3%; Score 28; DB 19; Length 354;
 Best Local Similarity 83.3%; Pred. No. 80;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKGK 6
 DB 5 mnvkgk 10

RESULT 9

R62508 R62508 standard; Protein; 359 AA.

AC R62508;

DT 26-JUN-1995 (first entry)

DE Galactosyl transferase clone product.

KM Gal-alpha (1,3) galactosyl transferase; xenograft; transplant;

XX rejection.

OS Sus scrofa domestica.

PN W09421799-A.

PD 29-SEP-1994.

PF 15-MAR-1994; 94WO-AU00126.

PR 16-MAR-1993; 93AU-0007854.

PA (AUST-) AUSTIN RES INST.

PI McKenzie IFC, Sandrin MS;

DR WPI: 1994-317019/39.

DR N-PSDB; Q74712.

PT DNA sequences encoding Gal-alpha (1,3)galactosyl transferase -
 PT and clones contg. such sequences are used in xenograft therapies

PS Disclosure: Page 35; 50pp; English.

XX The sequence is that of the product of the porcine Gal-alpha (1,3)
 CC galactosyl transferase gene which produces a Gal epitope on the
 CC surface of porcine cells. This epitope is recognised by antibodies
 CC which are responsible for hyperacute rejection of xenotransplanted
 CC pig cells, tissues and organs.
 CC See also R62507.

SO Sequence 359 AA:

Query Match 90.3%; Score 28; DB 15; Length 359;
 Best Local Similarity 83.3%; Pred. No. 81;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKGK 6
 DB 1 mnvkgk 6

RESULT 10

R90573 R90573 standard; Protein; 359 AA.

AC R90573;

DT 08-APR-1996 (first entry)

DE Pig alpha(1,3)-galactosyltransferase.

KM Alpha(1,3)galactosyltransferase; xenograft hyperacute rejection;
 KW transplantation; galactose alpha(1,3) galactose.

OS Sus scrofa.

PN W09534202-A1.

PD 21-DEC-1995.

PF 14-JUN-1995; 95WO-US07554.

PR 21-JUL-1994; 94US-0278282.

PR 15-JUN-1994; 94US-0260201.

PA (ALEX-) ALEXION PHARM INC.

PA (AUST-) AUSTIN RES INST.

PI Fodor WL, McKenzie IFC, Rother RP, Sandrin MS, Squinto SP;

DR WPI: 1996-049326/05.

DR N-PSDB; T12242.

PT Redn. of rejection of xenogeneic cells following transplantation
 PT by introducing a vector expressing fucosyl:transferase into the
 PT cells

PS Example 2; Page 52-54; 69pp; English.

XX Pig alpha(1,3)-galactosyltransferase (R90573) was expressed in
 CC monkey COS cells following transfection of the cells with
 CC vector pGT which contains an insert including the encoding
 CC cDNA (T12242). Co-transfection of these cells with vector pHT
 CC encoding human H-transferase (R90572) resulted in a reduction in
 CC the levels of galactose alpha(1,3) galactose epitopes expressed
 CC by the cells.

SO Sequence 359 AA:

Query Match 90.3%; Score 28; DB 17; Length 359;
 Best Local Similarity 83.3%; Pred. No. 81;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKGK 6
 DB 1 mnvkgk 6

RESULT 11

W49687 W49687 standard; Protein; 363 AA.

AC W49687;

DT 10-NOV-1998 (first entry)

DE Porcine alpha-1,3-galactosyl transferase isoform 2.

XX 13-APR-1994; 94US-0228933.
 XX (BIOT-) BIOTRANSPLANT INC.
 PA (GEHO) GEN HOSPITAL CORP.
 PA (CHIL-) INST CHILD HEALTH.
 XX Baetscher MW, Gustafsson KT, Sachs DH;
 XX WPI: 1995-373759/48.
 DR N-PSDB: T02892.
 XX Novel transgenic alpha (1,3) galactosyltransferase negative swine
 PT - used to produce rejection resistant cells for xenogenic
 PT transplantation
 PS Claim 11: Pages 35-37; 56pp; English.
 XX Transgenic swine in which the normal expression of the alpha (1,3)
 CC galactosyltransferase (AGT) R85082 is prevented, are prep'd. by
 CC inhibiting the expression of the AGT gene T02892 using antisense
 CC oligonucleotides or ribozyme inactivators in a pluripotent porcine
 CC embryonic stem cell. It is then inserted into a porcine oocyte
 CC (from which the pronuclear material has been removed), which is
 CC itself grown to produce the transgenic swine. Swine which do not
 CC express AGT will not produce carbohydrate moieties confg. the
 CC distinctive terminal Gal-alpha-1,3-Gal-beta-1-4GlcNAc epitope,
 CC which is a significant factor in xenogenic (esp. human) transplant
 CC rejection of swine grafts. Therefore the swine cells produced in
 CC the AGT negative transgenic swine are xenogenic transplant
 CC rejection resistant, and can therefore be used by a transplant
 CC recipient, or to provide gene therapy.
 XX Sequence 371 AA:
 SO
 CC This sequence represents isoform 2 of the porcine enzyme
 CC alpha-1,3-galactosyl transferase (alpha-1,3-GT). The enzyme catalyses
 CC the attachment of a galactose sugar molecule on the N-acetyllactosamine
 CC moiety found on surface glycoproteins and glycolipids. These sugar
 CC molecules are partly responsible for raising anti-graft antibodies, which
 CC lead to graft tissue rejection. The invention relates to a method of
 CC inhibiting the graft rejection mechanism by introducing the sequence
 CC encoding an antibody targeted to alpha-1,3-GT into the cells of animal,
 CC especially a pig, from whom organs may be used for xenotransplants.
 CC Neutralisation of the alpha-1,3-GT leads to tissues or organs lacking
 CC the galactose on the glycoproteins and glycolipids, thus preventing
 CC induction of the rejection response.
 CC Sequence 363 AA:
 SO
 Query Match 90.3%; Score 28; DB 19; Length 363;
 Best Local Similarity 83.3%; Pred. No. 82;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MNVKGK 6
 DB 5 MNVKGK 10
 RESULT 12
 R85082
 ID R85082 standard; Protein: 371 AA.
 AC R85082;
 XX
 DT 04-JUN-1996 (first entry)
 XX
 DE Porcine alpha (1,3) galactosyltransferase.
 XX
 XX Transgenic; swine; porcine; alpha (1,3) galactosyltransferase;
 KM antisense; ribozyme; Gal-alpha-1,3-Gal-beta-1-4GlcNAc; epitope;
 KM terminal; xenogenic; transplant; rejection; gene therapy; pig.
 XX
 OS Sus scrofa.
 XX
 PN WO9528412-A1.
 XX
 PD 26-OCT-1995.
 XX
 PF 21-MAR-1995; 95WO-US03940.

XX 13-APR-1994; 94US-0228933.
 XX (BIOT-) BIOTRANSPLANT INC.
 PA (GEHO) GEN HOSPITAL CORP.
 PA (CHIL-) INST CHILD HEALTH.
 XX Baetscher MW, Gustafsson KT, Sachs DH;
 XX WPI: 1995-373759/48.
 DR N-PSDB: T02892.
 XX Novel transgenic alpha (1,3) galactosyltransferase negative swine
 PT - used to produce rejection resistant cells for xenogenic
 PT transplantation
 PS Claim 11: Pages 35-37; 56pp; English.
 XX Transgenic swine in which the normal expression of the alpha (1,3)
 CC galactosyltransferase (AGT) R85082 is prevented, are prep'd. by
 CC inhibiting the expression of the AGT gene T02892 using antisense
 CC oligonucleotides or ribozyme inactivators in a pluripotent porcine
 CC embryonic stem cell. It is then inserted into a porcine oocyte
 CC (from which the pronuclear material has been removed), which is
 CC itself grown to produce the transgenic swine. Swine which do not
 CC express AGT will not produce carbohydrate moieties confg. the
 CC distinctive terminal Gal-alpha-1,3-Gal-beta-1-4GlcNAc epitope,
 CC which is a significant factor in xenogenic (esp. human) transplant
 CC rejection of swine grafts. Therefore the swine cells produced in
 CC the AGT negative transgenic swine are xenogenic transplant
 CC rejection resistant, and can therefore be used by a transplant
 CC recipient, or to provide gene therapy.
 XX Sequence 371 AA:
 SO
 Query Match 90.3%; Score 28; DB 16; Length 371;
 Best Local Similarity 83.3%; Pred. No. 84;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MNVKGK 6
 DB 1 MNVKGK 6
 RESULT 13
 R76777
 ID R76777 standard; Protein: 371 AA.
 AC R76777;
 XX
 DT 11-DEC-1995 (first entry)
 XX
 DE Pig alpha-1,3-galactosyltransferase.
 XX
 XX Alpha-1,3-galactosyltransferase; alpha-1,3-GalT; transgenic animal;
 KM pig; hyperacute rejection; xenotransplantation; donor organ;
 KM allograft rejection; Gal epitope; gene disruption;
 KM homologous recombination; knock-out.
 XX
 OS Sus scrofa.
 XX
 PN WO9520661-A1.
 XX
 PD 03-AUG-1995.
 XX
 PF 27-JAN-1995; 95WO-IB00088.
 XX
 XX 26-JAN-1995; 95US-0188607.
 PR 27-JAN-1994; 94US-0188607.
 XX
 PA (BRES-) BRESATEC LTD.
 PA (SVIN-) ST VINCENT'S HOSPITAL MELBOURNE LTD.

```

PI Crawford RJ, Dapice AJF, Pearce MJ, Rathjen PD:
PI Robbins AJ:
XX
XX MPI: 1995-275446/36.
DR N-PSDB: Q93077.
XX
PT New alpha-1,3-galactosyltransferase and leukaemia inhibitor factor
PT - corresp. DNA and nucleic acid constructs for inactivating the
PT transferase gene; for eliminating hyperacute region in human .
PT transplants
XX
XX
PS Claim 3; Fig.5; 184pp; English.
XX
CC cDNA encoding porcine alpha-1,3-GaIT was generated from liver RNA
CC using primers based on conserved regions of the mouse and cattle alpha-
CC 1,3-GaIT genes. Potential sites to interrupt the alpha-1,3-gaIt gene
CC (via homologous recombination) were identified in exons 4, 7, 8 and 9.
CC Such inactivation allows the breeding of 'knock-out' animals, e.g.
CC pigs suitable as donors of organs to overcome hyperacute rejection
CC problems in human xenotransplantation.
CC
SQ Sequence 371 AA;

Query Match 90.3%; Score 28; DB 16; Length 371;
Best Local Similarity 83.3%; Pred. No. 84;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0.

OY 1 MNVKGK 6
   |||||:
Db 1 mnvkgkr 6

RESULT 14
W49686
ID W49686 standard; Protein; 375 AA.
XX
XX W49686;
AC
XX
DT 10-NOV-1998 (first entry)
XX
DE Porcine alpha-1,3-galactosyl transferase isoform 1.
XX
XX Isoform: porcine; enzyme: alpha-1,3-galactosyl transferase; galactose;
KW sugar; N-acetyllactosamine; glycoprotein; glycolipid; antibody; pig;
KW graft tissue rejection; organ transplantation; xenotransplant.
XX
XX Sus scrofa.
OS
XX
XX FR2751346-A1.
FN
XX
XX 23-JAN-1998.
PD
XX
XX 19-JUL-1996; 96FR-0009077.
PF
XX
XX 19-JUL-1996; 96FR-0009077.
PR
XX
XX (INRM ) INSEEM INST NAT SANTE & RECH MEDICALE.
PA
PI Pourcel C, Soullion JP, Vanhove B:
XX
XX MPI: 1998-112876/11.
DR N-PSDB: V49453.
XX
XX Transgenic non-human donors of organs for human recipients -
PT containing DNA encoding antibodies that inhibit graft rejection
PT
XX
PS Claim 4; Page 32-34; 71pp; French.
XX
CC This sequence represents isoform 1 of the porcine enzyme
CC alpha-1,3-galactosyl transferase (alpha-1,3-GT). The enzyme catalyses
CC the attachment of a galactose sugar molecule on the N-acetyllactosamine

```

```

CC molecule found on surface glycoproteins and glycolipids. These sugar
CC molecules are partly responsible for raising anti-graft antibodies, which
CC lead to graft tissue rejection. The invention relates to a method of
CC inhibiting the graft rejection mechanism by introducing the sequence
CC encoding an antibody targeted to alpha-1,3-GT into the cells of animal,
CC especially a pig, from whom organs may be used for xenotransplants.
CC Neutralisation of the alpha-1,3-GT leads to tissues or organs lacking
CC the galactose on the glycoproteins and glycolipids, thus preventing
CC induction of the rejection response.
XX
XX Sequence 375 AA;
SQ
Query Match 90.3%: Score 28; DB 19; Length 375;
Best Local Similarity 83.3%: Pred. NO. 85;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 1 MNVKGK 6
|||||
Db 5 mnvkgk 10
RESULT 15
Y24083
ID Y24083 standard; peptide; 463 AA.
AC Y24083;
XX
XX 09-SEP-1999 (first entry)
XX
XX Salmonella typhimurium peptide sequence SEQ ID NO:52.
DE
XX
XX Salmonella: Salmonellosis; detection; diagnosis; infection; food.
XX
XX Salmonella typhimurium.
XX
XX OS
XX US5925522-A.
PN
XX
XX 20-JUL-1999.
PD
XX
XX 09-MAY-1997; 97US-0853659.
PF
XX
XX 09-MAY-1997; 97US-0853659.
PR
XX
XX (BAYT ) BATTLE MEMORIAL INST.
PA
XX
XX Saffer JD, Wong K;
PI
XX
XX WPI: 1999-418268/35.
DR
XX
XX N-PSDB: X83964.
XX
XX
XX Detecting Salmonella in humans, animals and food
XX
XX PT
XX
XX Disclosure: Column 187-190; 144pp: English.
XX
XX
XX The present sequence represents a Salmonella typhimurium peptide
XX sequence. The Salmonella nucleotide (X83935), its complement
XX (X83936) and fragments (X83937 to X83971), are useful for detecting
XX and diagnosing Salmonella infection in humans, animals and food. The
XX nucleotides will detect many, if not all Salmonella species especially
XX Salmonella dublin, S. enteritidis, S. gallinarum, S. minnesota,
XX S. paratyphi (types A, B and C), S. pullorum, S. typh tyza and
XX S. typhimurium. The fragments may be used to construct a DNA chip,
XX useful for the simple and inexpensive testing of substances for the
XX presence of Salmonella.
XX
XX Sequence 463 AA;
SQ
Query Match 90.3%: Score 28; DB 20; Length 463;
Best Local Similarity 83.3%: Pred. NO. 1,16+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

Oy 1 MNVKGK 6
|||:
Db 417 mnvrgk 422

Search completed: December 14, 2000, 10:11:10
Job time: 44 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December-14, 2000, 10:11:39 ; Search time 27.22 Seconds
(without alignments)
3.695 Million cell updates/sec

Title: US-09-051-034A-12

Perfect score: 31

Sequence: 1 MNVKGK 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 164575 seqs, 16761186 residues

Total number of hits satisfying chosen parameters: 164575

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/Backfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	100.0	368	2	US-08-378-617A-11
2	31	100.0	371	2	US-08-378-617A-12
3	31	100.0	376	2	US-08-704-548-2
4	31	100.0	394	1	US-07-914-281-4
5	31	100.0	394	1	US-08-393-246-4
6	31	100.0	394	1	US-08-525-058A-4
7	31	100.0	394	2	US-08-696-731-4
8	31	100.0	394	4	PCT-US91-00899-9
9	28	90.3	371	2	US-08-378-617A-10
10	28	90.3	463	2	US-08-853-659A-52
11	27	87.1	106	3	US-08-946-329A-76
12	27	87.1	187	1	US-08-441-629-11
13	27	87.1	187	3	US-08-776-207-11
14	27	87.1	187	4	PCT-US95-09172-11
15	27	87.1	429	4	US-08-307-896-4
16	27	87.1	429	4	PCT-US95-11808-4
17	27	87.1	533	3	US-08-726-214-18
18	27	87.1	1090	3	US-08-307-896-3
19	27	87.1	1090	3	US-08-726-214-4
20	27	87.1	1090	4	PCT-US95-11808-3
21	27	87.1	1099	4	US-08-726-214-14
22	27	87.1	1353	3	US-08-894-173-2
23	26	83.9	28	1	US-08-117-080-9
24	26	83.9	28	1	US-08-471-329-9
25	26	83.9	28	2	US-08-913-142-9
26	26	83.9	270	2	US-08-773-368-1
27	26	83.9	270	3	US-09-199-887-1
28	26	83.9	470	2	US-08-377-440A-1

ALIGNMENTS

29	25	80.6	87	3	US-09-023-082A-76	Sequence 76, App1
30	25	80.6	92	3	US-09-023-082A-82	Sequence 82, App1
31	25	80.6	106	3	US-09-023-082A-74	Sequence 74, App1
32	25	80.6	117	1	US-08-249-013-6	Sequence 6, App11
33	25	80.6	117	2	US-08-886-863-6	Sequence 6, App11
34	25	80.6	117	3	US-09-023-082A-72	Sequence 72, App1
35	25	80.6	117	4	PCT-US95-06764-6	Sequence 6, App11
36	25	80.6	119	3	US-09-023-082A-80	Sequence 80, App1
37	25	80.6	133	3	US-09-023-082A-70	Sequence 70, App1
38	25	80.6	140	3	US-09-023-082A-136	Sequence 136, App
39	25	80.6	141	3	US-09-023-082A-96	Sequence 96, App1
40	25	80.6	141	3	US-09-023-082A-112	Sequence 112, App
41	25	80.6	141	3	US-09-023-082A-120	Sequence 120, App
42	25	80.6	141	3	US-09-023-082A-124	Sequence 124, App
43	25	80.6	141	3	US-09-023-082A-128	Sequence 128, App
44	25	80.6	141	3	US-09-023-082A-132	Sequence 132, App
45	25	80.6	141	3	US-09-023-082A-140	Sequence 140, App

RESULT 1
US-08-378-617A-11
; Sequence 11, Application US/08378617A
; Patent No. 5849991
; GENERAL INFORMATION:
; APPLICANT: d'Apice, Anthony J.F.
; APPLICANT: Pearce, Martin J.
; APPLICANT: Robins, Allan J.
; APPLICANT: Crawford, Robert J.
; APPLICANT: Rathjen, Peter D.
; TITLE OF INVENTION: MATERIALS AND METHODS FOR MANAGEMENT OF
; TITLE OF INVENTION: HYPERACUTE REJECTION IN HUMAN XENOTRANSPLANTATION
; NUMBER OF SEQUENCES: 33
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 120 South Sixth Street, Suite 2500
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/378.617A
; FILING DATE: 26-JAN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Ellinger, Mark S.
; REGISTRATION NUMBER: 34,812
; REFERENCE/DOCKET NUMBER: 06868/005001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (612) 335-5070
; TELEFAX: (612) 288-9696
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 368 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-378-617A-11

Query Match 100.0%; Score 31; DB 2; Length 368;
Best Local Similarity 100.0%; Pred. No. 20;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNVKGK 6
||||||

Db 1 MNVKGK 6

RESULT 2

US-08-378-617A-12

Sequence 12, Application US/08378617A

Patent No. 5849991

GENERAL INFORMATION:

APPLICANT: d'Apice, Anthony J.F.

APPLICANT: Pearce, Martin J.

APPLICANT: Robins, Allan J.

APPLICANT: Crawford, Robert J.

APPLICANT: Rathjen, Peter D.

TITLE OF INVENTION: MATERIALS AND METHODS FOR MANAGEMENT OF

NUMBER OF SEQUENCES: 33

CORRESPONDENCE ADDRESS:

ADDRESSEE: Fish & Richardson

STREET: 120 South Sixth Street, Suite 2500

CITY: Minneapolis

STATE: MN

COUNTRY: USA

ZIP: 55402

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30B

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/378,617A

FILING DATE: 26-JAN-1995

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Ellinger, Mark S.

REGISTRATION NUMBER: 34,812

TELECOMMUNICATION INFORMATION:

TELEPHONE: (612) 335-5070

TELEFAX: (612) 288-9696

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 371 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-378-617A-12

Query Match 100.0%; Score 31; DB 2; Length 371;

Best Local Similarity 100.0%; Pred. No. 20;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKGK 6

Db 1 MNVKGK 6

RESULT 3

US-08-704-548-2

Sequence 2, Application US/08704548

Patent No. 5879675

GENERAL INFORMATION:

APPLICANT: GALILI, URI

APPLICANT: REPIC, PATRICIA M.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR VACCINES

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: SEIDEL, GONDA, LAVORGNA & MONACO, P.C.

STREET: Suite 1800, Two Penn Center Plaza

CITY: Philadelphia

STATE: PA

COUNTRY: USA

ZIP: 19102

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/704,548

FILING DATE: 11-SEP-1996

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Monaco, Daniel A.

REGISTRATION NUMBER: 30,480

REFERENCE/DOCKET NUMBER: 8760-2 C11

TELECOMMUNICATION INFORMATION:

TELEPHONE: (215) 568-8383

TELEFAX: (215) 568-5549

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 376 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-704-548-2

Query Match 100.0%; Score 31; DB 2; Length 376;

Best Local Similarity 100.0%; Pred. No. 21;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKGK 6

Db 1 MNVKGK 6

RESULT 4

US-07-914-281-4

Sequence 4, Application US/07914281

Patent No. 5324663

GENERAL INFORMATION:

APPLICANT: LOWE, JOHN B.

TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS

TITLE OF INVENTION: OF OLISACCHARIDE STRUCTURES ON GLYCOPROTEINS,

TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,

ADDRESS: P.C.

STREET: 1755 Jefferson Davis Highway, Fourth Floor

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/914,281

FILING DATE: 19920720

CLASSIFICATION: 530

ATTORNEY/AGENT INFORMATION:

NAME: Lavalleye, Jean-Paul M. P.

REGISTRATION NUMBER: 31,451

REFERENCE/DOCKET NUMBER: 2363-060-55

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703)521-4500

TELEFAX: (703)486-2347

TELEX: 248855 OPAT UR

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 394 amino acids
TYPE: AMINO ACID
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-07-914-281-4

Query Match 100.0%; Score 31; DB 1; Length 394;

Best Local Similarity 100.0%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MNVKGK 6
Db 36 MNVKGK 41

RESULT 5
US-08-393-246-4
Sequence 4, Application US/08393246
Patent No. 5595900

GENERAL INFORMATION:
APPLICANT: LOWE, JOHN B.
TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESS: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/393,246
FILING DATE:
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/220,433
FILING DATE: 30-MAR-1994
APPLICATION NUMBER: US 07/914,281
FILING DATE: 20-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Lavalleye, Jean-Paul M. P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-060-55
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 394 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-393-246-4

Query Match 100.0%; Score 31; DB 1; Length 394;

Best Local Similarity 100.0%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MNVKGK 6
Db 36 MNVKGK 41

RESULT 6
US-08-525-058A-4
Sequence 4, Application US/08525058A
Patent No. 5770420

GENERAL INFORMATION:
APPLICANT: LOWE, JOHN B.
TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTURES
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESS: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/525,058A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lavalleye, Jean-Paul M. P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-060-55
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 394 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-525-058A-4

Query Match 100.0%; Score 31; DB 1; Length 394;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MNVKGK 6
Db 36 MNVKGK 41

RESULT 7
US-08-696-731-4
Sequence 4, Application US/08696731
Patent No. 5955347

GENERAL INFORMATION:
APPLICANT: LOWE, JOHN B.
TITLE OF INVENTION: METHODS AND PRODUCTS FOR THE SYNTHESIS
TITLE OF INVENTION: OF OLIGOSACCHARIDE STRUCTURES ON GLYCOPROTEINS,
TITLE OF INVENTION: GLYCOLIPIDS, OR AS FREE MOLECULES, AND FOR THE ISOLATION
TITLE OF INVENTION: OF CLONED GENETIC SEQUENCES THAT DETERMINE THESE STRUCTU
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESS: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
STREET: 1755 Jefferson Davis Highway, Fourth Floor
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.

Query Match 100.0%; Score 31; DB 1; Length 394;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/696,731
FILING DATE: 14-AUG-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/393,246
FILING DATE:
APPLICATION NUMBER: US 08/220,433
FILING DATE: 30-MAR-1994
APPLICATION NUMBER: US 07/914,281
FILING DATE: 20-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Lavalleye, Jean-Paul M. P.
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-060-55
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-4500
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 394 amino acids
TYPE: amino acid
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-696-731-4

Query Match 100.0%; Score 31; DB 2; Length 394;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNVKGK 6
|||||
Db 36 MNVKGK 41

RESULT 8
PCT-US91-00899-9
Sequence 9, Application PC/RTUS9100899
GENERAL INFORMATION:
APPLICANT: Lowe, John B.
TITLE OF INVENTION: Method and Products For the Synthesis of
TITLE OF INVENTION: Oligosaccharide Structures on Glycoproteins, Glycolipids,
or as Free Molecules, and for the Isolation of Cloned
TITLE OF INVENTION: Genetic Sequences That Determine These Structur
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,
ADDRESSEE: P.C.
STREET: 1755 Jefferson Davis Highway, Suite 400
CITY: Arlington
STATE: Virginia
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US91/00899
FILING DATE: 19910214
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Lavalleye Ph.D., Jean-Paul
REGISTRATION NUMBER: 31,451
REFERENCE/DOCKET NUMBER: 2363-021-55 PCT

TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)521-5940
TELEFAX: (703)486-2347
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 394 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: YES
ORIGINAL SOURCE: Blood
TISSUE TYPE: Blood
PCT-US91-00899-9

Query Match 100.0%; Score 31; DB 4; Length 394;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNVKGK 6
|||||
Db 36 MNVKGK 41

RESULT 9
US-08-378-617A-10
Sequence 10, Application US/08378617A
Patent No. 5849991
GENERAL INFORMATION:
APPLICANT: d'Alpe, Anthony J.F.
APPLICANT: Pearse, Martin J.
APPLICANT: Robins, Allan J.
APPLICANT: Crawford, Robert J.
APPLICANT: Rathjen, Peter D.
TITLE OF INVENTION: MATERIALS AND METHODS FOR MANAGEMENT OF
TITLE OF INVENTION: HYPERACUTE REJECTION IN HUMAN XENOTRANSPLANTATION
NUMBER OF SEQUENCES: 33
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson
STREET: 120 South Sixth Street, Suite 2500
CITY: Minneapolis
STATE: MN
COUNTRY: USA
ZIP: 55402
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30B
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/378,617A
FILING DATE: 26-JAN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Ellinger, Mark S.
REGISTRATION NUMBER: 34,812
REFERENCE/DOCKET NUMBER: 06668/005001
TELECOMMUNICATION INFORMATION:
TELEPHONE: (612) 335-5070
TELEFAX: (612) 288-9696
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 371 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-378-617A-10

Query Match 90.3%; Score 28; DB 2; Length 371;
Best Local Similarity 83.3%; Pred. No. 87;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKGK 6
|||||

Db 1 MNVKGK 6

RESULT 10
US-08-853-659A-52
; Sequence 52, Application US/088536559A
; Patent No. 5925322
; GENERAL INFORMATION:
; APPLICANT: Wong, K.K.; Saffar, J.D.
; TITLE OF INVENTION: A Salmonella Sequence, Methods of Detection
; TITLE OF INVENTION: Of A
; TITLE OF INVENTION: Salmonella Sequence, And Methods of Detection of Salmonella
; NUMBER OF SEQUENCES: 67
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Paul W. Zimmerman
; ADDRESSEE: Intellectual Property Services
; ADDRESSEE: Battelle Memorial Institute
; ADDRESSEE: P.O. Box 999
; STREET: Washington Way
; CITY: Richland
; STATE: Washington
; COUNTRY: U.S.A.
; ZIP: 99352
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette 3.50", 1.44 MB storage
; COMPUTER: IBM PC/XT/AT
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: Word Processor (WordPerfect 5.1)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/853.659A
; FILING DATE: Unknown
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: none
; FILING DATE: n/a
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 463 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; US-08-853-659A-52

Query Match 90.3%; Score 28; DB 2; Length 463;
Best Local Similarity 83.3%; Pred. No. 1,1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKGK 6
|||||

Db 417 MNVKGK 422

RESULT 11
US-08-946-329A-76
; Sequence 76, Application US/08946329A
; Patent No. 6057091
; GENERAL INFORMATION:
; APPLICANT: Beechey, Phillip A.
; APPLICANT: Porter, Jeffrey A.
; TITLE OF INVENTION: NOVEL HEDGEHOG-DERIVED POLYPEPTIDES
; NUMBER OF SEQUENCES: 109
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: USA
; ZIP: 92037
; COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows 95
; SOFTWARE: FASTSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/946.329A
; FILING DATE: 07-OCT-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/061,323
; FILING DATE: 07-OCT-1996
; APPLICATION NUMBER: 08/729,743
; FILING DATE: 10-JUL-1996
; APPLICATION NUMBER: 08/567,357
; FILING DATE: 04-DEC-1995
; APPLICATION NUMBER: 08/349,498
; FILING DATE: 02-DEC-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Hallie, Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 07265/140001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619/678-5070
; TELEFAX: 619/678-5099
; INFORMATION FOR SEQ ID NO: 76:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 106 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-946-329A-76

Query Match 87.1%; Score 27; DB 3; Length 106;
Best Local Similarity 83.3%; Pred. No. 43;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKGK 6
|||||

Db 8 INVKGK 13

RESULT 12
US-08-441-629-11
; Sequence 11, Application US/08441629
; Patent No. 5766923
; GENERAL INFORMATION:
; APPLICANT: Kirschner, Marc W.
; APPLICANT: Kinoshita, No. 57669231yuk1
; TITLE OF INVENTION: RECEPTOR-LIGAND ASSAY
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Millitia Drive
; CITY: Lexington
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/441.629
; FILING DATE: 15-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/279,217
; FILING DATE: 22-JUL-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Granahan, Patricia
; REGISTRATION NUMBER: 32,227
; REFERENCE/DOCKET NUMBER: HU95-01A

```
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
FAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 187 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-441-629-11

Query Match      87.1% Score 27; DB 1; Length 187;
Best Local Similarity 83.3% Pred. No. 73;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNVKGK 6
Db 120 MNAKGK 125

RESULT 13
US-08-776-207-11
Sequence 11, Application US/08776207A
Patent No. 6080718
GENERAL INFORMATION:
APPLICANT: Kirschner, Marc W.
APPLICANT: Kirschner, No. 6080718yuki
TITLE OF INVENTION: Receptor-Ligand Assay
FILE REFERENCE: HU95-01A2
CURRENT APPLICATION NUMBER: US/08/776,207A
CURRENT FILING DATE: 1997-06-23
EARLIER APPLICATION NUMBER: PCT/US95/09172
EARLIER FILING DATE: 1995-07-19
EARLIER APPLICATION NUMBER: 08/441,629
EARLIER FILING DATE: 1995-05-15
EARLIER APPLICATION NUMBER: 08/279,217
EARLIER FILING DATE: 1994-07-22
NUMBER OF SEQ ID NOS: 18
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 11
LENGTH: 187
TYPE: PRT
ORGANISM: Xenopus laevis
US-08-776-207-11

Query Match      87.1% Score 27; DB 3; Length 187;
Best Local Similarity 83.3% Pred. No. 73;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNVKGK 6
Db 120 MNAKGK 125

RESULT 14
PCT-US95-09172-11
Sequence 11, Application PC/TUS9509172
GENERAL INFORMATION:
APPLICANT: Kirschner, Marc W.
APPLICANT: Kirschner, No. 6080718yuki
TITLE OF INVENTION: RECEPTOR-LIGAND ASSAY
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESS: Hamilton, Brook, Smith & Reynolds, P.C.
STREET: Two Millitia Drive
CITY: Lexington
STATE: Massachusetts
COUNTRY: USA
ZIP: 02173
COMPUTER-READABLE FORM:
```

```
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/09172
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/279,217
FILING DATE: 22-JUL-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/441,629
FILING DATE: 15-MAY-1995
ATTORNEY/AGENT INFORMATION:
NAME: Granahan, Patricia
REGISTRATION NUMBER: 32,227
REFERENCE/DOCKET NUMBER: HU95-01A PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 861-6240
FAX: (617) 861-9540
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 187 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US95-09172-11
```

```
Query Match      87.1% Score 27; DB 4; Length 187;
Best Local Similarity 83.3% Pred. No. 73;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNVKGK 6
Db 120 MNAKGK 125

RESULT 15
US-08-307-896-4
Sequence 4, Application US/08307896C
Patent No. 6034071
GENERAL INFORMATION:
APPLICANT: Iyengar, Sriivas Ravi
TITLE OF INVENTION: MUTANT ACTIVATED GSALPHA AND ADENYLYL
FILE REFERENCE: 29770
CURRENT APPLICATION NUMBER: US/08/307,896C
CURRENT FILING DATE: 1994-09-16
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 429
TYPE: PRT
ORGANISM: Homo sapiens
US-08-307-896-4

Query Match      87.1% Score 27; DB 3; Length 429;
Best Local Similarity 83.3% Pred. No. 1.6e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNVKGK 6
Db 401 INVKGK 406
```

Search completed: December 14, 2000, 10:11:40
Job time: 74 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OW protein - protein search, using sw model

Run on: December 14, 2000, 10:12:56 ; Search time 27.57 seconds
(without alignments)
13.811 Million cell updates/sec

Title: US-09-051-034A-12

Perfect score: 31

Sequence: 1 MNVKGK 6

Scoring table:

BL0SUM62
Gapop 10.0 , Gapext 0.5

Searched: 182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters: 182106

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31	100.0	26	A39769	N-acetylglucosamin
2	31	100.0	368	A44785	N-acetylglucosamin
3	31	100.0	371	I49698	alpha-1,3-galactos
4	31	100.0	376	A56480	N-acetylglucosamin
5	31	100.0	394	A34417	alpha-1,3-mannosyl
6	30	96.8	386	BLRC	leucine/isoleucine
7	28	90.3	213	F71804	probable transcrip
8	28	90.3	367	S14619	branched-chain aml
9	28	90.3	369	A71025	hypothetical prote
10	28	90.3	371	I46583	alpha-1,3-galactos
11	28	90.3	378	T30965	hypothetical prote
12	28	90.3	416	K1HOG	phosphoglycerate k
13	28	90.3	458	C72591	hypothetical prote
14	28	90.3	463	T14884	hypothetical prote
15	28	90.3	1166	T13958	synap-b1 protein
16	28	90.3	1249	T14270	Ras-GRase activat
17	28	90.3	1293	T14259	ras GRase-activat
18	27	87.1	98	G71671	ribosomal protein
19	27	87.1	187	S23595	embryonic fibrobla
20	27	87.1	192	S54407	embryonic fibrobla
21	27	87.1	347	T23889	hypothetical prote
22	27	87.1	348	T11013	MLC536.28c protei
23	27	87.1	390	H69483	hypothetical prote
24	27	87.1	418	G70558	hypothetical prote
25	27	87.1	428	T19309	hypothetical prote
26	27	87.1	454	S61052	hypothetical prote
27	27	87.1	473	A25040	levansucrase (EC 2
28	27	87.1	476	I37136	adenylate cyclase
29	27	87.1	511	T11940	ribosomal protein

30	27	87.1	567	2	C71963	hypothetical prote
31	27	87.1	602	2	T03057	hypothetical prote
32	27	87.1	639	2	C42049	leishmanolysin (EC
33	27	87.1	694	2	A69768	transcription anti
34	27	87.1	812	2	E71972	pyruvate, water dik
35	27	87.1	812	2	A64535	pyruvate, water dik
36	27	87.1	869	2	S72760	psp1 protein - Myc
37	27	87.1	1090	2	A41541	adenylate cyclase
38	27	87.1	1099	2	A55405	adenylate cyclase
39	27	87.1	1353	2	JC4279	adenylate cyclase
40	27	87.1	1395	2	S25997	gene atpA intron 1
41	27	87.1	1451	2	S42167	190K protein - hum
42	27	87.1	2340	2	B71704	cell surface anti
43	27	87.1	4919	2	T31105	hypothetical prote
44	26	83.9	44	2	JC6154	viral capsid prote
45	26	83.9	95	1	R5BS23	ribosomal protein

ALIGNMENTS

RESULT 1
A39769
N-acetylglucosaminide alpha-1,3-galactosyltransferase (EC 2.4.1.151) homologue - human
C:Species: Homo sapiens (man)
C:Date: 14-Feb-1992 #sequence_revision 14-Feb-1992 #text_change 13-Sep-1998
C:Accession: A39769
R:Joziasse, D.H.; Shaper, J.H.; Jabs, E.W.; Shaper, N.L.
J. Biol. Chem. 266, 6891-6998, 1991
A:Title: Characterization of an alpha1-->3-galactosyltransferase homologue on human c
A:Reference number: A39769; MUID:91201351
A:Accession: A39769
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-26 <J02>
A:Superfamily: histo-blood group 1 transferase
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 100.0%; Score 31; DB 2; Length 26;
Best Local Similarity 100.0%; Pred. No. 1.2;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKGK 6
Db 1 MNVKGK 6

RESULT 2
A44785
N-acetylglucosamine 3-alpha-galactosyltransferase (EC 2.4.1.124) - bovine
C:Species: Bos primigenius taurus (cattle)
C:Date: 19-Mar-1993 #sequence_revision 19-Mar-1993 #text_change 23-Sep-1999
C:Accession: A44785
J:Rozaasae, D.H.; Shaper, J.H.; Van den Eljnden, D.H.; Van Tunen, A.J.; Shaper, N.L.
J. Biol. Chem. 264, 14290-14297, 1989
A:Title: Bovine alpha1->3-galactosyltransferase: Isolation and characterization of a
A:Reference number: A44785; MUID:89340543
A:Accession: A44785
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-368 <J02>
A:Cross-references: GB:J04989; NID:G163123; PIDN:AAA30558.1; PID:G163124
C:Superfamily: histo-blood group 1 transferase
C:Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein

Query Match 100.0%; Score 31; DB 2; Length 368;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MNVKGK 6
||||||

Db 1 MNVKGK 6

RESULT 3

149698

alpha-1,3-galactosyltransferase - mouse

A:Species: Mus musculus (house mouse)

C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 29-Sep-1999

C:Accession: 149698

R:Jozkisse, D.H.; Shaper, N.L.; Kim, D.; van der Eijnden, D.H.; Shaper, J.H.

J. Biol. Chem. 267, 5534-5541, 1992

A:Title: Murine alpha-1,3-galactosyltransferase: A single gene locus specifies four iso-

A:Reference number: A42117; MUID:92184813

A:Accession: 149698

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-371 <RES>

A:Cross-references: GB:M85153; NID:g193563; PIDN:AAA37711.1; PID:g457142

A:Genetics:

A:Gene: Ggta-1

C:Superfamily: histo-blood group 1 transferase

Query Match

Best Local Similarity 100.0%; Score 31; DB 2; Length 371;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKGK 6

Db 1 MNVKGK 6

RESULT 4

A56480

N-acetylglucosaminide alpha-1,3-galactosyltransferase (EC 2.4.1.151) - marmoset

C:Species: Callithrix sp.

C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 29-Sep-1999

C:Accession: A56480

R:Henion, T.R.; Macher, B.A.; Anarakli, F.; Gallili, U.

Glycobiology 4, 193-201, 1994

A:Title: Defining the minimal size of catalytically active primate alpha1,3 galactosyltr

A:Reference number: A56480; MUID:94331837

A:Accession: A56480

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-376 <HEN>

A:Cross-references: GB:S71333; NID:g558051; PIDN:AA831587.1; PID:g558052

A:Note: authors translated the codon GTG for residue 251 as Ser

C:Superfamily: histo-blood group 1 transferase

C:Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein

Query Match 100.0%; Score 31; DB 2; Length 376;

Best Local Similarity 100.0%; Pred. No. 14;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKGK 6

Db 1 MNVKGK 6

RESULT 5

A34417

alpha-1,3-mannosyl-glycoprotein beta-1,4-N-acetylglucosaminyltransferase (EC 2.4.1.145)

C:Species: Mus musculus (house mouse)

C:Date: 15-Jun-1990 #sequence_revision 15-Jun-1990 #text_change 29-Sep-1999

C:Accession: A34417

R:Larsen, R.D.; Rajan, V.P.; Ruff, M.M.; Kukowska-Latallo, J.; Cummings, R.D.; Lowe, J.B.

Proc. Natl. Acad. Sci. U.S.A. 86, 8227-8231, 1989

A:Title: Isolation of a cDNA encoding a murine UDPgalactose:beta-D-galactosyl-1,4-N-acet

A:Reference number: A34417; MUID:90046765

A:Accession: A34417

A:Status: preliminary

A:Molecule type: mRNA
A:Residues: 1-394 <LAR>
A:Cross-references: GB:M26925; NID:g193419; PIDN:AAA37657.1; PID:g309242
C:Superfamily: histo-blood group 1 transferase
C:Keywords: glycosyltransferase; hexosyltransferase; transmembrane protein

Query Match 100.0%; Score 31; DB 2; Length 394;
Best Local Similarity 100.0%; Pred. No. 15;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKGK 6

Db 36 MNVKGK 41

RESULT 6

BLEC

leucine/isoleucine/valine-binding protein precursor - Escherichia coli

N:Alternate names: leucine transport protein livJ precursor; LIV-binding protein

C:Species: Escherichia coli

C:Date: 31-May-1979 #sequence_revision 05-Dec-1997 #text_change 16-Jul-1999

C:Accession: G65142; A23576; A37074; A03415; S47679; I55524

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.

A.: Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617

A:Accession: G65142

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-386 <BLAT>

A:Cross-references: GB:AE000422; GB:U00096; NID:g1789868; PIDN:AAC76485.1; PID:g17898

A:Experimental source: strain K-12, substrain MG1655

R:Landick, R.; Oxender, D.L.

J. Biol. Chem. 260, 8257-8261, 1985

A:Title: The complete nucleotide sequences of the Escherichia coli LIV-BP and LS-BP g

A:Reference number: A94677; MUID:85234531

A:Accession: A23576

A:Molecule type: DNA

A:Residues: 20-21, 'T', 23-30, 'L', 32-88, 'A', 90-386 <LAN>

A:Cross-references: GB:J05516; GB:K02178; GB:M10426; GB:M13166; NID:g14663

R:Adams, M.D.; Wagner, L.M.; Graddis, T.J.; Landick, R.; Antonucci, T.K.; Gibson, A.L.

J. Biol. Chem. 265, 11436-11443, 1990

A:Title: Nucleotide sequence and genetic characterization reveal six essential genes

A:Reference number: A37074; MUID:90307651

A:Accession: A37074

A:Molecule type: DNA

A:Residues: 20-21, 'T', 23-30, 'L', 32-88, 'A', 90-386 <ADA>

A:Cross-references: GB:J05516; NID:g146630; PIDN:AA83881.1; PID:g146631

R:Ovchinnikov, V.A.; Aldanova, N.A.; Grinkevich, V.A.; Arzamazova, N.M.; Moroz, I.N.

FEBS Lett. 78, 313-316, 1977

A:Title: The primary structure of a leu, ile and val (LIV)-binding protein from Esche

A:Reference number: A03415; MUID:77225636

A:Accession: A03415

A:Molecule type: protein

A:Residues: 43-304, 'AD', 307-383, 'AD', 386 <OVC>

A:Experimental source: strain K12

A:Note: this protein was isolated from the periplasmic space of the cell

R:Plunkett, G.

submitted to the EMBL Data Library, March 1994

A:Reference number: S47666

A:Accession: S47679

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-386 <PLU>

A:Cross-references: EMBL:U00039; NID:g466582; PIDN:AA818435.1; PID:g912457

A:Note: this reported protein translation starts with a GTG codon; it has an extra 23

R:Antonucci, T.K.; Landick, R.; Oxender, D.L.

J. Cell. Biochem. 29, 209-216, 1985

A:Title: The leucine binding proteins of Escherichia coli as models for studying the
A:Reference number: I55524; MUID:86066153
A:Accession: I55524

A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 20-21,'T','23-30','L','32-88','A','90-386 <RES>
A:Cross-references: GB:M29377; NID:g416628; PIDN:AAA24075.1; PID:g416629
C:Genetics:
A:Gene: livJ
A:Map position: 76 min
C:Function:
A:Description: specifically binds Leu, Ile, and Val and may be involved in their transme
C:Superfamily: LIV-binding protein
F:1-42/Domain: (or 20-42) signal sequence #status predicted <SIG>
F:43-386/Product: leucine/isoleucine/valine-binding protein #status experimental <MAT>
F:95-120/Dissulfide bonds: #status experimental

Query Match 96.88; Score 30; DB 1; Length 386;
Best Local Similarity 83.38; Pred. No. 25;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKGK 6
||:||||
Db 20 MNVKGK 25

RESULT 7
F71804
probable transcription regulator - Helicobacter pylori (strain J99)
C:Species: Helicobacter pylori
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence-revision 12-Feb-1999 #text-change 08-Oct-1999
C:Accession: F71804; A71826
R:Alm, R.A.; Ling, L.S.L.; Moir, D.T.; King, B.L.; Brown, E.D.; Dolg, P.C.; Smith, D.R.;
Mature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; MUID:99120557
A:Accession: F71804
A:Molecule type: DNA
A:Residues: 1-213 <ARN>
A:Cross-references: GB:AE001567; GB:AE001439; NID:g4156065; PIDN:AAD07036.1; PID:g415608
A:Experimental source: strain J99
A:Accession: A71826
A:Molecule type: DNA
A:Residues: 1-213 <ARN>
A:Cross-references: GB:AE001551; GB:AE001439; NID:g4155887; PIDN:AAD06868.1; PID:g416500
A:Experimental source: strain J99
C:Genetics:
A:Gene: jhp1443
A:Note: similar to H. pylori 26695 gene HP1365
C:Superfamily: ompr protein; response regulator homology
C:Keywords: phosphoprotein
F:5-115/Domain: response regulator homology <RRH>
F:53/Binding site: phosphate (ASP) (covalent) #status predicted

Query Match 90.38; Score 28; DB 2; Length 213;
Best Local Similarity 83.38; Pred. No. 41;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKGK 6
||:||||
Db 137 LNVKKG 142

RESULT 8
S14619
branched-chain amino acid-binding protein livJ - Citrobacter freundii
N:Alternate names: LIV-binding protein livJ
C:Species: Citrobacter freundii
C:Date: 20-Feb-1995 #sequence-revision 20-Feb-1995 #text-change 20-Aug-1999
C:Accession: S14619
R:Dagget Garvin, L.; Hardie, S.C.
submitted to the EMBL Data Library, April 1991

A:Description: Nucleotide sequence of the livJ gene from Citrobacter freundii.
A:Reference number: S14619
A:Accession: S14619
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-367 <DAG>
A:Cross-references: EMBL:X58820; NID:g40470; PIDN:CAA1622.1; PID:g40471
C:Superfamily: LIV-binding protein

Query Match 90.38; Score 28; DB 2; Length 367;
Best Local Similarity 83.38; Pred. No. 68;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKGK 6
||:||||
Db 1 MNVKGK 6

RESULT 9
A71025
hypothetical protein PH1494 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence-revision 14-Aug-1998 #text-change 20-Jun-2000
C:Accession: A71025
R:Kawabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Se
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kishida, N.; Ogu
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophil
A:Reference number: A71000; MUID:98344137
A:Accession: A71025
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-369 <XAM>
A:Cross-references: GB:AP000006; NID:g3236133; PIDN:BAH30601.1; PID:g3257918
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by Genba
C:Genetics:
A:Gene: PH1494
C:Superfamily: Pyrococcus horikoshii hypothetical protein PH1494

Query Match 90.38; Score 28; DB 2; Length 369;
Best Local Similarity 83.38; Pred. No. 68;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKGK 6
||:||||
Db 216 LNVKKG 221

RESULT 10
I46583
alpha-1,3-galactosyltransferase - pig
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 21-Feb-1997 #sequence-revision 21-Feb-1997 #text-change 29-Sep-1999
C:Accession: I46583
R:Strahan, K.M.; Gu, F.; Preece, A.F.; Gustavsson, I.; Andersson, L.; Gustafsson, K.
Immunogenetics 41, 101-105, 1995
A:Title: cDNA sequence and chromosome localization of pig alpha 1,3 galactosyltransfe
A:Reference number: I46583; MUID:95104914
A:Accession: I46583
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-371 <STR>
A:Cross-references: GB:U36152; NID:g642635; PIDN:AAA73558.1; PID:g642636
C:Genetics:
A:Gene: GGTAL
C:Superfamily: histo-blood group 1 transferase

Query Match 90.38; Score 28; DB 2; Length 371;
Best Local Similarity 83.38; Pred. No. 69;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 OY 1 MNVKGK 6
 |||||
 Db 1 MNVKGK 6

RESULT 11

T30965.1
 hypochelical protein R04B3.2 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 21-Jan-2000
 C:Accession: T30965
 R:Gething, S.; Pauley, A.
 A:Submitted to the EMBL Data Library, August 1999
 A:Description: The sequence of C. elegans cosmid R04B3.
 A:Reference number: Z20947
 A:Accession: T30965
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-378 <GAT>
 A:Cross-references: EMBL:U50198; PIDN:AAA91260.1
 A:Experimental source: strain Bristol N2
 C:Genetics:
 A:Map position: X
 A:Insertions: 42/1; 182/3; 222/2; 251/3; 322/1
 A:Notes: R04B3.2
 C:Superfamily: N4-(beta-N-acetylglucosaminyl)-L-asparaginase

Query Match 90.3%; Score 28; DB 2; Length 378;
 Best Local Similarity 83.3%; Pred. No. 70;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKGK 6
 |||||
 Db 1 MNVKGK 332

RESULT 12

K1H003.1
 phosphoglycerate kinase (EC 2.7.2.3) - horse
 C:Species: Equus caballus (domestic horse)
 C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 28-Feb-1997
 C:Accession: A92292; A00669
 R:Merrett, M.
 J. Biol. Chem. 256, 10293-10305, 1981
 A:Title: Primary structure of 3-phosphoglycerate kinase from horse muscle. II. Amino acid complete sequence of the enzyme.
 A:Reference number: A92292; MUID:82030789
 A:Accession: A92292
 A:Molecule type: protein
 A:Residues: 1-416 <MEB>
 R:Bank, R.D.; Blake, C.C.F.; Evans, P.R.; Haser, R.; Rice, D.W.; Hardy, G.W.; Merrett, Nature 279, 773-777, 1979
 A:Title: Sequence, structure and activity of phosphoglycerate kinase: a possible hinge-hinge structure.
 A:Reference number: A93209; MUID:79199779
 A:Accession: A93209
 A:Contents: annotation; X-ray crystallography; 2.5 angstroms; muscle
 C:Comment: The structure consists of two discrete, globular domains that are joined by a 405-406 form a helix associated with the amino-terminal domain.
 C:Comment: Residues thought to be involved in ADP-ATP binding are Glu-343 to ribose and 39-341
 C:Superfamily: phosphoglycerate kinase
 C:Key words: acetylated amino end; ATP; gluconeogenesis; glycolysis; phosphotransferase P1/ATP-binding site: acetylated amino end (Ser) #status experimental
 F:219-283/Binding site: ATP (Lys, Glu) #status experimental

Query Match 90.3%; Score 28; DB 1; Length 416;
 Best Local Similarity 83.3%; Pred. No. 76;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKGK 6

Db 11 LNVKGR 16
 :|||||

RESULT 13

C72591
 hypothetical protein APE1199 - Aeropyrum pernix (strain K1)
 C:Species: Aeropyrum pernix
 C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
 C:Accession: C72591
 R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Halkawa, Y.; Jin-no, K.; Tanaka, H.; Takamiya, M.; Masuda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J. DNA Res. 6, 83-101, 1999
 A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aero
 A:Reference number: A72450; MUID:99310339
 A:Accession: C72591
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-458 <KAW>
 A:Cross-references: DBJ:AP000061; NID:95104821; PIDN:BA80185.1; PID:95104871
 A:Experimental source: strain K1
 C:Genetics:
 A:Gene: APE1199

Query Match 90.3%; Score 28; DB 2; Length 458;
 Best Local Similarity 83.3%; Pred. No. 83;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKGK 6
 :|||||
 Db 31 LNVKGR 36

RESULT 14

T14884
 hypothetical protein - Salmonella typhimurium
 C:Species: Salmonella typhimurium
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T14884
 R:Wong, K.K.; McClelland, M.; Stillwell, L.C.; Sisk, E.C.; Thurston, S.J.; Saffer, J. Infect. Immun. 66, 3365-3371, 1998
 A:Title: Identification and sequence analysis of a 27-kilobase chromosomal fragment c
 A:Reference number: T14884
 A:Accession: T14884
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-463 <WON>
 A:Cross-references: EMBL:AF060869; NID:93323584; PID:93323601; PIDN:ACC26653.1
 C:Genetics:
 A:Gene: spl4_P
 A:Map position: 92 min

Query Match 90.3%; Score 28; DB 2; Length 463;
 Best Local Similarity 83.3%; Pred. No. 84;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKGK 6
 :|||||
 Db 417 MNVKGK 422

RESULT 15

T13958
 synGAP-b1 protein - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
 C:Accession: T13958
 R:Suzuki, T.
 submitted to the EMBL Data Library, August 1998
 A:Description: SynGAP-b1.

A:Reference number: 217834
 A:Accession: T13958
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1166 <SUZ>
 A:Cross-references: EMBL:AB016962; NID:d1261311; PID:d1038706; PIDN:CAB19493.1
 A:Experimental source: strain Sprague Dawley
 C:Genetics:
 A:Gene: syncAP-b1

Query Match 90.3%; Score 28; DB 2; Length 1166;
 Best Local Similarity 83.3%; Pred. No. 2e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 MNVKGK 6
 :|||||
 Db 322 LNVKGK 327

Search completed: December 14, 2000, 10:12:57
 Job time: 150 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 14, 2000, 10:14:57 ; Search time 15.82 seconds

(without alignments)
12.117 Million cell updates/sec

Title: US-09-051-034a-12

Sequence: 31
1 MNVKGK 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters: 87993

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31	100.0	368	1	GATR_BOVIN
2	31	100.0	394	1	GATR_MOUSE
3	28	90.3	367	1	LIVJ_CITFR
4	28	90.3	371	1	GATR_PIG
5	28	90.3	378	1	ASPG_CAEEL
6	28	90.3	416	1	PGK_HORSE
7	28	90.3	749	1	PEX_HUMAN
8	28	90.3	749	1	PEX_MOUSE
9	27	87.1	98	1	RL23_RICPR
10	27	87.1	187	1	FGFB_XENLA
11	27	87.1	192	1	FGFB_XENLA
12	27	87.1	367	1	LIVJ_ECOLI
13	27	87.1	473	1	SACB_BACSU
14	27	87.1	476	1	CVA2_HUMAN
15	27	87.1	511	1	RT04_PROVI
16	27	87.1	812	1	PPSA_HELPI
17	27	87.1	812	1	PPSA_HELPI
18	27	87.1	869	1	YEBI_MOUSE
19	27	87.1	1080	1	CVA7_HUMAN
20	27	87.1	1090	1	CVA2_RAT
21	27	87.1	1099	1	CVA7_MOUSE
22	27	87.1	1353	1	CVA9_HUMAN
23	27	87.1	1353	1	CVA9_MOUSE
24	27	87.1	1451	1	MYML_HUMAN
25	26	83.9	95	1	RL23_BACST
26	26	83.9	95	1	RL23_BACST
27	26	83.9	96	1	RL23_BACST
28	26	83.9	184	1	VP33_RHORD
29	26	83.9	259	1	VP33_NPVAC
30	26	83.9	273	1	HMDI_CAEEL
31	26	83.9	282	1	VP33_NPVOP
32	26	83.9	319	1	Y184_MYCPN
33	26	83.9	375	1	Y785_METJA

34	26	83.9	389	1	TRB1_THEMA	P50909 thermologa
35	26	83.9	519	1	YMP4_YEAST	Q04347 saccharomyc
36	26	83.9	647	1	HS70_XENLA	P02827 xenopus lae
37	26	83.9	874	1	CPHA_SYNY8	P56947 synectocyst
38	26	83.9	1357	1	YJ03_YEAST	P47104 saccharomyc
39	26	83.9	1934	1	MYSB_MESAU	P13340 mesocricetu
40	26	83.9	1938	1	MYSA_MOUSE	Q02566 mus musculu
41	26	83.9	1938	1	MYSA_RAT	P02566 mus musculu
42	26	83.9	1939	1	MYSA_MESAU	P13539 mesocricetu
43	25	80.6	82	1	Y409_RICPR	Q92483 rickettsia
44	25	80.6	141	1	Y0FB_CAEEL	Q09338 caenorhabd
45	25	80.6	194	1	FGF4_CHICK	P48804 gallus gall

ALIGNMENTS

RESULT ID	1	GATR_BOVIN	STANDARD:	PRT:	368 AA.
AC	P14769:				
DT	01-APR-1990 (rel. 14, Created)				
DT	01-APR-1990 (rel. 14, Last sequence update)				
DT	30-MAY-2000 (rel. 39, Last annotation update)				
DE	N-ACETYLGLYCOSAMINYL-1,3-GALACTOSYLTRANSFERASE (EC 2.4.1.151)				
DE	(GALACTOSYLTRANSFERASE) (UDP-GALACTOSE:BETA-D-GALACTOSYL-1,4-N-ACETYL-				
DE	D-GALACTOSAMINYL-1,3-GALACTOSYLTRANSFERASE).				
GN	CGTAL.				
OS	Bos taurus (Bovine).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;				
OC	Bovidae; Bovinae; Bos.				
RP	[1]				
RN	SEQUENCE FROM N.A.				
RX	MEDLINE: 89340543.				
RA	Joelasse D.H., Shaper J.H., van den Eljnden D.H., van Tunen A.J.,				
RA	Shaper N.L.;				
RT	*Bovine alpha 1->3-galactosyltransferase: Isolation and				
RT	characterization of a cDNA clone. Identification of homologous				
RT	sequences in human genomic DNA.";				
RL	J. Biol. Chem. 264:14290-14297(1989).				
CC	- FUNCTION: TRANSFER OF GALACTOSE FROM UDP-GALACTOSE TO AN				
CC	ACCEPTOR MOLECULE (R).				
CC	- CATALYTIC ACTIVITY: UDP-GALACTOSE + BETA-D-GALACTOSYL-1,4-				
CC	N-ACETYL-D-GALACTOSAMINYL-R = UDP + ALPHA-D-GALACTOSYL-1,3-				
CC	BETA-D-GALACTOSYL-1,4-N-ACETYL-D-GALACTOSAMINYL-R.				
CC	- COFACTOR: MANGANESE.				
CC	- PATHWAY: GLYCOSYLATION.				
CC	- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND				
CC	FORM IN TRANS CISTERNAE OF GOLGI.				
CC	- DISEASE: AUTOIMMUNE DISEASE (ANTIBODIES AGAINST ENZYMATIC				
CC	PRODUCT).				
CC	- SIMILARITY: STRONG, TO BLOOD GROUP AB TRANSFERASE.				
CC	- SIMILARITY: STRUCTURAL SIMILARITY WITH THE OTHER MAMMALIAN				
CC	GLYCOSYLTRANSFERASES.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation				
CC	at the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch)				
CC	-----				
DR	EMBL: J04989; AAA30558.1; ..				
DR	PIR: A44785; A44785.				
KW	Transferase; Glycosyltransferase; Glycoprotein; Transmembrane;				
KW	Signal-anchor; Golgi stack; Manganese.				
FT	DOMAIN 1 6				
FT	TRANSMEM 22				
FT	DOMAIN 23 368				
FT	CARDIOPHYD 293 293				
SQ	SEQUENCE 368 AA: 43246 MW: 585506737BDDC33 CRC64:				

AC P50127; 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE N-ACETYLGLUCOSAMINIDE ALPHA-1,3-GALACTOSYLTRANSFERASE (EC 2.4.1.151)
DE (GALACTOSYLTRANSFERASE) (UDP-GALACTOSE:BETA-D-GALACTOSYL-1,4-N-ACETYL-
D-GALUCOSAMINIDE ALPHA-1,3-GALACTOSYLTRANSFERASE).
GN GGTAL.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-YORKSHIRE: TISSUE-SPLEEN;
MEDLINE: 95104914.
RA Strahan K.M., Gu F., Preece A.F., Gustavsson I., Andersson L.,
Gustafsson K.;
RT cDNA sequence and chromosome localization of pig alpha 1,3
galactosyltransferase.";
RL Immunogenetics 41:101-105(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE-SPLEEN, AND LIVER;
RA Sendin M.S., Dabkowski P.L., Henning M.M., Moutouris E.,
McKenzie I.F.C.;
RL Submitted (FEB-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: TRANSFER OF GALACTOSE FROM UDP-GALACTOSE TO AN
ACCEPTOR MOLECULE (R).
CC -1- CATALYTIC ACTIVITY: UDP-GALACTOSE + BETA-D-GALACTOSYL-1,4-
N-ACETYL-D-GALUCOSAMINYL-R = UDP + ALPHA-D-GALACTOSYL-1,3-
BETA-D-GALACTOSYL-1,4-N-ACETYL-D-GALUCOSAMINYL-R.
CC -1- COFACTOR: MANGANESE.
CC -1- PATHWAY: GLYCOSYLATION.
CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. MEMBRANE-BOUND
FORM IN TRANS CISTERNAE OF GOLGI.
CC -1- SIMILARITY: STRONG, TO BLOOD GROUP AB TRANSFERASE.
CC -1- SIMILARITY: STRUCTURAL SIMILARITY WITH THE OTHER MAMMALIAN
GLYCOSYLTRANSFERASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L36153; AAA7358.1; -
DR EMBL: L36535; AAA58775.1; -
KW Transferrase; Glycosyltransferase; Glycoprotein; Transmembrane;
KW Signal-anchor; Golgi stack; Alternative splicing; Manganese.
FT DOMAIN 1 6
FT TRANSMEM 1 22
FT DOMAIN 23 371
FT DOMAIN 59 59
FT CARBOHYD 296 296
FT CARBOHYD 27 39
FT VARSPLIC 227 227
FT CONFLICT 227 227
FT SEQUENCE 371 AA; 43764 MW; CFC715EBD8993D4 CRC64;
Query Match 90.3%; Score 28; DB 1; Length 371;
Best Local Similarity 83.3%; Pred. No. 26;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 1 MNVKGK 6
DB 1 MNVAGR 6
RESULT 5
ASPG_CAEEL STANDARD; PRT; 378 AA.
ID ASPG_CAEEL

AC Q21697; 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PUTATIVE N4-(BETA-N-ACETYLGLUCOSAMINYL)-L-ASPARAGINASE PRECURSOR
DE (EC 3.5.1.26) (GLYCOSYLASPARAGINASE) (ASPARTYLGLUCOSAMINIDASE) (N4-(N-
ACETYL-BETA-GLUCOSAMINYL)-L-ASPARAGINE AMIDASE) (AGA).
GN R04B3.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Pauley A., Gattung S.;
RL Submitted (MAR-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: CLEAVES THE GLCNAC-ASN BOND WHICH JOINS OLIGOSACCHARIDES
TO THE PEPTIDE OF ASPARAGINE-LINKED GLYCOPROTEINS (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: N4-(BETA-N-ACETYL-D-GLUCOSAMINYL)-L-ASPARAGINE
+ H(2)O = N-ACETYL-BETA-GLUCOSAMINYLAMINE + L-ASPARTATE.
CC -1- SUBCELLULAR LOCATION: LYSOSOMAL (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE ASPARAGINASE 2 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U50198; AAA91260.1; -
DR HSSP: P20933; 1APZ.
DR WORMPEP: R04B3.2; CE07412.
DR INTERPRO: IPR000246; -
DR PFAM: PF01112; Asparaginase_2; 1.
KW Hypothetical protein; Glycoprotein; Hydrolase; Lysosome; signal.
FT SIGNAL 1 19
FT CHAIN 20 226
FT CHAIN 227 378
FT CHAIN 62 67
FT DISULFID 176 192
FT DISULFID 337 361
FT ACT_SITE 227 227
FT CARBOHYD 15 15
FT CARBOHYD 133 133
FT CARBOHYD 211 211
FT CARBOHYD 351 351
FT SEQUENCE 378 AA; 41012 MW; C9694965167DC6A CRC64;
Query Match 90.3%; Score 28; DB 1; Length 378;
Best Local Similarity 83.3%; Pred. No. 26;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 1 MNVKGK 6
DB 327 MNVAGR 332
RESULT 6
PGK_HORSE STANDARD; PRT; 416 AA.
ID PGK_HORSE
AC P00559;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE PHOSPHOGLYCERATE KINASE (EC 2.7.2.3).
GN PGK.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;

OC Mammalia; Euthera; Perissodactyla; Equidae; Equus.
 RN [1]
 RN SEQUENCE.
 RX MEDLINE: 82030789.
 RA McCrell M.;
 RT "Primary structure of 3-phosphoglycerate kinase from horse muscle.
 RT 11. Amino acid sequence of cyanogen bromide peptides CBI-CB4 and
 RT CB6-CB14, sequence of methionine-containing regions, and complete
 RT sequence of the enzyme.";
 RL J. Biol. Chem. 256:10293-10305(1981).
 RN [2]
 RN X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS).
 RC TISSUE=MUSCLE;
 RX MEDLINE: 79199779.
 RA Banks R.D., Blake C.C.F., Evans P.R., Haser R., Rice D.W.,
 RA Hardy G.W., Merrett M., Phillips A.W.;
 RT "Sequence, structure and activity of phosphoglycerate kinase: a
 RT possible hinge-bending enzyme.";
 RL Nature 279:773-777(1979).
 CC -1- CATALYTIC ACTIVITY: ATP + 3-PHOSPHO-D-GLYCERATE = ADP +
 CC 3-PHOSPHO-D-GLYCEROL PHOSPHATE.
 CC -2- PATHWAY: SECOND STEP IN THE SECOND PHASE OF GLYCOLYSIS.
 CC -3- SUBUNIT: MONOMER.
 CC -4- SIMILARITY: BELONGS TO THE PHOSPHOGLYCERATE KINASE FAMILY.
 DR PDB: A92292; KIHOG.
 DR PDB: 2PGK; 17-FEB-84.
 DR INTERPRO: IPR001576; .
 DR Pfam: PF00162; PGK; 1.
 DR PRINTS: PR00477; PHGICXKINASE.
 DR PROSITE: PS00111; GLYCERATE_KINASE; 1.
 KW Transferase; Kinase; Acetylation; Glycolysis; 3D-structure.
 FT MOD_RES 1 185
 FT DOMAIN 1 185
 FT DOMAIN 186 189
 FT DOMAIN 190 416
 FT DOMAIN 405 416
 FT BINDING 219 219
 FT AGC_SITE 212 214
 FT AGC_SITE 236 238
 FT AGC_SITE 339 341
 FT BINDING 343 343
 FT SEQUENCE 416 AA; 44471 MW; A4B926A6FE0B0FDA CRC64;
 Query Match 90.3%; Score 28; DB 1; Length 416;
 Best Local Similarity 83.3%; Pred. No. 29;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 MNVKGK 6
 Db 11 LNVKKG 16
 RESULT 7
 PEX_HUMAN STANDARD; PRT; 749 AA.
 ID P07562; 000678; 099827; 013646; 093032;
 AC 01-NOV-1997 (Rel. 35; Created)
 DT 01-NOV-1997 (Rel. 35; Last sequence update)
 DT 15-JUL-1998 (Rel. 36; Last annotation update)
 DE PHOSPHATE REGULATING NEUTRAL ENDOPEPTIDASE (EC 3.4.24.-)
 DE (METALLOENDOPEPTIDASE HOMOLOG PEX) (X-LINKED HYPOPHOSPHATEMIA PROTEIN)
 DE (HBP) (VITAMIN D-RESISTANT HYPOPHOSPHATEMIC RICKETS PROTEIN).
 GN PEX.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Euthera; Primates; Catarrhini; Homiidae; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Francis F., Strom T.M., Hennig S., Boeddrich A., Lorenz B.,
 RA Brandau O., Mohnike K.L., Cagnoli M., Steffens C., Klages S.,
 RA Borzym K., Pohl T., Oudet C., Econs M.J., Rowe P.S.N., Reinhardt R.,
 RA Meitinger T., Lehrach H.;

RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RN SEQUENCE FROM N.A.
 RX MEDLINE: 97232252.
 RA Beck L., Soumounou Y., Martel J., Krishnamurthy G., Gauthier C.,
 RA Goodyer C.G., Tenenhouse H.S.;
 RT "Pex/PEX tissue distribution and evidence for a deletion in the 3'
 RT region of the pex gene in X-linked hypophosphatemic mice.";
 RL J. Clin. Invest. 99:1200-1209(1997).
 RN [3]
 RN SEQUENCE FROM N.A.
 RC TISSUE=BONE;
 RX MEDLINE: 97343443.
 RA Guo R., Quarles L.D.;
 RT "Cloning and sequencing of human PEX from a bone cDNA library:
 RT evidence for its developmental stage-specific regulation in
 RT osteoblasts.";
 RL J. Bone Miner. Res. 12:1009-1017(1997).
 RN [4]
 RN SEQUENCE FROM N.A.
 RP Lipman M.L., Panda D., Henderson J.E., Shen Y., Goltzman D.,
 RA Karaplis A.C.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RN SEQUENCE FROM N.A.
 RX MEDLINE: 97224400.
 RA Gileff M., Mumm S., Maelitz P., Mazzarella R., Whyte M.P.,
 RA Thakker R.V., Schlesinger D.;
 RT "Expression and cloning of the human X-linked hypophosphatemia gene
 RT cDNA.";
 RL Biochem. Biophys. Res. Commun. 231:635-639(1997).
 RN [6]
 RN SEQUENCE OF 4-641 FROM N.A.
 RP MEDLINE: 96024647.
 RA Francis F., Hennig S., Korn B., Reinhardt R., de Jong P.,
 RA Poustka A., Lehrach H., Rowe P.S.N., Goulding J.N., Pronicka E.,
 RA Summerfield T., Mountford R., Read A.P., Popowska E.,
 RA Davies K.E., Orlordan J.L.H., Econs M.J., Nesbitt T.,
 RA Drezer M.K., Oudet C., Pannetier S., Hanauer A., Strom T.M.,
 RA Meindl A., Lorenz B., Cagnoli M., Mohnike K.L., Murken J.,
 RA Meitinger T.;
 RT "A gene (PEX) with homologues to endopeptidases is mutated in
 RT patients with X-linked hypophosphatemic rickets. The Hyp
 RT Consortium.";
 RL Nat. Genet. 11:130-136(1995).
 RN [7]
 RN VARIANTS HYP TYR-85; CYS-166; SER-252; ILE-253 AND VAL-579.
 RX MEDLINE: 97260404.
 RA Holm I.A., Huang X., Kunkel L.M.;
 RT "Mutational analysis of the pex gene in patients with X-linked
 RT hypophosphatemic rickets.";
 RL Am. J. Hum. Genet. 60:790-797(1997).
 RN [8]
 RN VARIANTS HYP SER-77; PRO-138; LEU-534 AND ARG-579.
 RP MEDLINE: 97252387.
 RA Rowe P.S.N., Oudet C.L., Francis F., Sindling C., Pannetier S.,
 RA Econs M.J., Strom T.M., Meitinger T., Garabedian M., David A.,
 RA Machon M.-A., Questiaux E., Popowska E., Pronicka E., Read A.P.,
 RA Mokrzycki A., Glorieux F.H., Drezer M.K., Hanauer A., Lehrach H.,
 RA Goulding J.N., Orlordan J.L.H.;
 RT "Distribution of mutations in the pex gene in families with X-linked
 RT hypophosphatemic rickets (HYP).";
 RL Hum. Mol. Genet. 6:539-549(1997).
 CC -1- FUNCTION: PROBABLY INVOLVED IN BONE AND DENTIN MINERALIZATION
 CC AND RENAL PHOSPHATE REABSORPTION.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).
 CC -1- TISSUE SPECIFICITY: LYMPHOCYTE AND FETAL BRAIN; NOT IN ADULT
 CC BRAIN, PLACENTA, SKELETAL MUSCLE, AND PANCREAS; NOT IN ADULT AND
 CC FETAL HEART, LUNG, LIVER, AND KIDNEY.
 CC -1- DISEASE: DEFECTS IN PEX ARE A CAUSE OF X-LINKED HYPOPHOSPHATEMIC
 CC RICKETS (HYP), A DOMINANT DISORDER CHARACTERIZED BY IMPAIRED
 CC PHOSPHATE UPTAKE IN THE KIDNEY, WHICH IS LIKELY TO BE CAUSED BY
 CC ABNORMAL REGULATION OF SODIUM PHOSPHATE COTRANSPORT IN THE

CC PROXIMAL TUBULES.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M13 (ZINC
 CC METALLOPROTEASE). ALSO KNOWN AS THE NEPRILYSIN SUBFAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X10196: CAA71258.1: -
 DR EMBL: U75645: AAB47749.1: -
 DR EMBL: U82970: AAC24487.1: -
 DR EMBL: U87284: AAB47352.1: -
 DR EMBL: Y08111: CAA69326.1: -
 DR EMBL: Y08112: CAA69326.1: JOINED.
 DR EMBL: Y08113: CAA69326.1: JOINED.
 DR EMBL: Y08114: CAA69326.1: JOINED.
 DR EMBL: Y08115: CAA69326.1: JOINED.
 DR EMBL: Y08116: CAA69326.1: JOINED.
 DR EMBL: Y08117: CAA69326.1: JOINED.
 DR EMBL: Y08118: CAA69326.1: JOINED.
 DR EMBL: Y08119: CAA69326.1: JOINED.
 DR EMBL: Y08120: CAA69326.1: JOINED.
 DR EMBL: Y08121: CAA69326.1: JOINED.
 DR EMBL: Y08122: CAA69326.1: JOINED.
 DR EMBL: Y08123: CAA69326.1: JOINED.
 DR EMBL: Y08124: CAA69326.1: JOINED.
 DR EMBL: Y08125: CAA69326.1: JOINED.
 DR EMBL: Y08126: CAA69326.1: JOINED.
 DR EMBL: Y08127: CAA69326.1: JOINED.
 DR EMBL: Y08128: CAA69326.1: JOINED.
 DR EMBL: Y08129: CAA69326.1: JOINED.
 DR EMBL: Y08130: CAA69326.1: JOINED.
 DR EMBL: Y08132: CAA69326.1: JOINED.
 DR EMBL: Y08183: AAB42219.1: -
 DR EMBL: Y08187: AAB42219.1: JOINED.
 DR EMBL: Y08168: AAB42219.1: JOINED.
 DR EMBL: Y08169: AAB42219.1: JOINED.
 DR EMBL: Y08171: AAB42219.1: JOINED.
 DR EMBL: Y08172: AAB42219.1: JOINED.
 DR EMBL: Y08173: AAB42219.1: JOINED.
 DR EMBL: Y08174: AAB42219.1: JOINED.
 DR EMBL: Y08175: AAB42219.1: JOINED.
 DR EMBL: Y08176: AAB42219.1: JOINED.
 DR EMBL: Y08177: AAB42219.1: JOINED.
 DR EMBL: Y08178: AAB42219.1: JOINED.
 DR EMBL: Y08179: AAB42219.1: JOINED.
 DR EMBL: Y08180: AAB42219.1: JOINED.
 DR EMBL: Y08661: AAB42219.1: JOINED.
 DR EMBL: Y08181: AAB42219.1: JOINED.
 DR EMBL: Y08182: AAB42219.1: JOINED.
 DR EMBL: Y0800712: AAB51604.1: -
 DR EMBL: Y080475: AAC50552.1: -
 DR MIM: 307800: -
 DR INTERPRO: IPR000130: -
 DR INTERPRO: IPR000718: -
 DR PFAM: PF01431: Peptidase_M13: 1.
 DR PRINTS: PR00786: NEPRILYSIN.
 DR PROSITE: PS00142: ZINC_PROTEASE_1
 KW Hydrolase; Metalloprotease; Amino-peptidase; Zinc; Glycoprotein;
 KW Transmembrane; Signal-anchor; Disease mutation.
 FT DOMAIN 1 20 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 21 41 SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 FT (POTENTIAL).
 FT DOMAIN 42 641 EXTRACELLULAR (POTENTIAL).
 FT METAL 580 580 ZINC (CATALYTIC) (BY SIMILARITY).
 FT ACT_SITE 581 581 BY SIMILARITY.
 FT METAL 584 584 ZINC (CATALYTIC) (BY SIMILARITY).
 FT METAL 642 642 ZINC (CATALYTIC) (BY SIMILARITY).

FT	ACT_SITE	646	646	PROTON DONOR (POTENTIAL).
FT	CARBOHYD	71	71	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	238	238	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	263	263	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	280	280	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	301	301	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	377	377	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	484	484	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	736	736	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	VARIANT	77	77	C -> S (IN HYP).
FT	VARIANT	85	85	/FTID-VAR_006738.
FT	VARIANT	138	138	/FTID-VAR_006739.
FT	VARIANT	166	166	L -> P (IN HYP).
FT	VARIANT	166	166	/FTID-VAR_006740.
FT	VARIANT	252	252	R -> C (IN HYP).
FT	VARIANT	252	252	/FTID-VAR_006741.
FT	VARIANT	253	253	F -> S (IN HYP).
FT	VARIANT	253	253	/FTID-VAR_006742.
FT	VARIANT	534	534	M -> I (IN HYP).
FT	VARIANT	534	534	/FTID-VAR_006743.
FT	VARIANT	579	579	P -> L (IN HYP).
FT	VARIANT	579	579	/FTID-VAR_006744.
FT	VARIANT	579	579	G -> R (IN HYP).
FT	VARIANT	579	579	/FTID-VAR_006745.
FT	VARIANT	579	579	G -> V (IN HYP).
FT	CONFLICT	363	363	/FTID-VAR_006746.
FT	CONFLICT	403	403	A -> D (IN REF. 6).
FT	CONFLICT	641	641	W -> R (IN REF. 6).
FT	CONFLICT	641	641	G -> A (IN REF. 6).
FT	SEQUENCE	749 AA:	86473 MW:	7C4F9F3E2471C6A8 CRC64:

Query Match 90.3%; Score 28; DB 1; Length 749;
 Best Local Similarity 83.3%; Pred. No. 51;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNVKG 6

Db 632 LNVKG 637

RESULT 8
 PEX_MOUSE STANDARD: PRT: 749 AA.
 AC P70699; P97439; 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE METALLOENDOPEPTIDASE HOMOLOG PEX (EC 3.4.24.-) (PHOSPHATE REGULATING
 DE NEUTRAL ENDOPEPTIDASE) (X-LINKED HYPHOPHATHEMIA PROTEIN) (HYP
 DE (VITAMIN D-RESISTANT HYPHOPHATHEMIC RICKETS PROTEIN).
 GN PEX OR PEXH OR HYP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 96411643.
 RA Du L., Desparats M., Viel J., Giorieux F.H., Cawthorn C., Ecarot B.,
 RA "CDNA cloning of the murine Pex gene implicated in X-linked
 RT hypophosphatemia and evidence for expression in bone.";
 RL Genomics 36:22-28(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 97217775.
 RA Strom T.M., Francis F., Lorenz B., Boeddrich A., Econs M.J.,
 RA Lehrach H., Meltinger T.,
 RT "Pex gene deletions in Gy and Hyp mice provide mouse models for
 RT X-linked hypophosphatemia.";
 RL Hum. Mol. Genet. 6:165-171(1997).
 RN [3]
 RP SEQUENCE FROM N.A.

CC STRAIN=C57BL/6J;
 RX MEDLINE; 97232252.
 RA Beck L., Soumounou Y., Martel J., Krishnamurthy G., Gauthier C.,
 RA Geodier C.G., Temenhoun H.S.;
 RT "Pex/PEX tissue distribution and evidence for a deletion in the 3'
 RT region of the Pex gene in X-linked hypophosphatemic mice";
 RL J. Clin. Invest. 99:1200-1209(1997).
 CC -1- FUNCTION: PROBABLY INVOLVED IN BONE AND DENTIN MINERALIZATION
 CC AND RENAL PHOSPHATE REABSORPTION.
 CC -1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN (POTENTIAL).
 CC -1- TISSUE SPECIFICITY: BONE.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M13 (ZINC
 CC METALLOPROTEASE); ALSO KNOWN AS THE NEPHRILYSIN SUBFAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U49508; AAC36502.1; -
 CC EMBL; U73910; AAC25962.1; -
 CC EMBL; U73912; AAC25964.1; -
 CC EMBL; U73913; AAC25965.1; -
 CC EMBL; U73914; AAC25966.1; -
 CC EMBL; U73915; AAC25967.1; -
 CC EMBL; U73915; AAC25967.1; -
 CC EMBL; U75646; AAB47750.1; -
 CC MGD; MGI:107489; PHEX.
 CC INTERPRO: IPR000130; -
 CC INTERPRO: IPR000718; -
 CC PFM; PF01431; Peptidase_M13; 1.
 CC BRENTS; PRO0786; NEPHRILYSIN.
 CC PROSITE; PS00142; ZINC_PROTEASE; 1.
 CC K0 Hydroxylase; Metalloprotease; Zinc; Glycoprotein; Transmembrane;
 CC Signal-anchor.
 CC
 CC TRANSMEM 1 20 CYTOPLASMIC (POTENTIAL).
 CC SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
 CC (POTENTIAL).
 CC
 CC BOWAIN 38 749 EXTRACELLULAR (POTENTIAL).
 CC METAL 580 580 ZINC (CATALYTIC) (BY SIMILARITY).
 CC ACET-SITE 581 581 BY SIMILARITY.
 CC METAL 584 584 ZINC (CATALYTIC) (BY SIMILARITY).
 CC METAL 642 642 ZINC (CATALYTIC) (BY SIMILARITY).
 CC ACET-SITE 646 646 PROTON DONOR (POTENTIAL).
 CC CARBOHYD 71 71 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 263 263 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 301 301 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 377 377 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 484 484 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC CARBOHYD 736 736 N-LINKED (GLCNAC. . .) (POTENTIAL).
 CC GMPFLUCT 430 430 D -> V (IN REF. 3).
 CC SEQUENCE 749 AA; 86418 MW; A88FA481C376C18A CRC64;

Query Match 90.3%; Score 28; DB 1; Length 749;
 Best Local Similarity 83.3%; Pred. No. 51;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKGK 6
 DB 632 LNVKKG 637

RESULT 9
 RL23_RICPR STANDARD: PRT: 98 AA.
 AC 092C07;
 DT 30-MAY-2000 (Rel. 39, Created)

DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE 50S RIBOSOMAL PROTEIN L23.
 GN RPLM OR RP657.
 OS Rickettsia prowazekii.
 CC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 CC Rickettsiaceae; Rickettsiae; Rickettsia.
 CC [1]
 CC SEQUENCE FROM N.A.
 CC STRAIN=MADRID E;
 CC MEDLINE; 99039499.
 CC
 CC Andersson S.G.E., Zomrodipour A., Andersson J.O.,
 CC Sichenitz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,
 CC Eriksson A.-S., Winkler H.H., Kurland C.G.;
 CC "The genome sequence of Rickettsia prowazekii and the origin of
 CC mitochondria";
 CC Nature 396:133-140(1998).
 CC
 CC -1- FUNCTION: BINDS TO A SPECIFIC REGION ON THE 23S rRNA
 CC (BY SIMILARITY).
 CC
 CC -1- SIMILARITY: BELONGS TO THE L23P FAMILY OF RIBOSOMAL PROTEINS.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AJ235272; CAA15097.1; -
 CC INTERPRO: IPR001014; -
 CC PFM; PF00276; Ribosomal_L23; 1.
 CC PROSITE; PS00050; RIBOSOMAL_L23; FALSE_NEG.
 CC K0 Ribosomal protein; rRNA-binding.
 CC SEQUENCE 98 AA; 11411 MW; 22F5448F3DE2BBD0F CRC64;

Query Match 87.1%; Score 27; DB 1; Length 98;
 Best Local Similarity 66.7%; Pred. No. 12;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKGK 6
 DB 59 LNVKKG 64

RESULT 10
 FGFA_XENLA STANDARD: PRT: 187 AA.
 AC P48805;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 01-FEB-1996 (Rel. 33, Last annotation update)
 DE FIBROBLAST GROWTH FACTOR-4-1 PRECURSOR (FGF-4-1) (EMBRYONIC
 DE FIBROBLAST GROWTH FACTOR 1) (XFGF-1).
 DE Xenopus laevis (African clawed frog).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
 CC Xenopodinae; Xenopus.
 CC [1]
 CC SEQUENCE FROM N.A.
 CC MEDLINE; 92315916.
 CC Isaacs H.V., Tannahill D., Slack J.M.W.;
 CC "Expression of a novel FGF in the Xenopus embryo. A new candidate
 CC inducing factor for mesoderm formation and anteroposterior
 CC specification";
 CC Development 114:711-720(1992).
 CC
 CC -1- FUNCTION: GOOD CANDIDATE FOR AN INDUCING FACTOR WITH POSSIBLE
 CC ROLES BOTH IN MESODERM INDUCTION AT THE BLASTULA STAGE AND IN THE
 CC FORMATION OF THE ANTEROPOSTERIOR AXIS AT THE GASTRULA STAGE.
 CC -1- SUBCELLULAR LOCATION: SECRETED (PROBABLE).
 CC -1- SIMILARITY: BELONGS TO THE HEPARIN-BINDING GROWTH FACTORS FAMILY.
 CC
 CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X62593; CAA44479.1; -
DR HSSP: P09038; 2BRH.
DR INTERPRO: IPR002209; -
DR PFAM: PF00167; FGF; 1.
DR PRINTS: PR00262; ILIHGFG.
DR PROSITE: PS00247; HBGFG_FGF; 1.
KW Growth factor; Mitogen; Signal.
FT SIGNAL
FT CHAIN
SQ SEQUENCE 187 AA; 21223 MW; AAE63D65E82AD1BD CRC64;

Query Match 87.1%; Score 27; DB 1; Length 187;
Best Local Similarity 83.3%; Pred. No. 23;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNVKGK 6
DB 120 MNVKGK 125

RESULT 11
RQFB_XENLA STANDARD; PRT; 192 AA.
ID FGFB_XENLA
AC P48806;
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE FIBROBLAST GROWTH FACTOR-4-II PRECURSOR (FGF-4-II) (HBGF-4-II)
DE (EMBRONIC FIBROBLAST GROWTH FACTOR II) (XERGF-II).
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92315916.
RA Isaacs H.V., Tannahill D., Slack J.M.W.;
RT "Expression of a novel FGF in the Xenopus embryo. A new candidate
RT inducing factor for mesoderm formation and anteroposterior
RT specification";
RT Development 114:711-720(1992).
CC -1- FUNCTION: GOOD CANDIDATE FOR AN INDUCING FACTOR WITH POSSIBLE
CC ROLES BOTH IN MESODERM INDUCTION AT THE BLASTULA STAGE AND IN THE
CC FORMATION OF THE ANTEROPOSTERIOR AXIS AT THE GASTRULA STAGE.
CC -1- SUBCELLULAR LOCATION: SECRETED (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE HEPARIN-BINDING GROWTH FACTORS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X62594; CAA44480.1; -
DR HSSP: P09038; 1BFF.
DR INTERPRO: IPR002209; -
DR PFAM: PF00167; FGF; 1.
DR PRINTS: PR00262; ILIHGFG.
DR PROSITE: PS00263; HBGFGFG.

DR PROSITE: PS00247; HBGFG_FGF; 1.
KW Growth factor; Mitogen; Signal.
FT SIGNAL
FT CHAIN
SQ SEQUENCE 192 AA; 21903 MW; 2B01B0B8824E3B3 CRC64;

Query Match 87.1%; Score 27; DB 1; Length 192;
Best Local Similarity 83.3%; Pred. No. 23;
Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNVKGK 6
DB 125 MNVKGK 130

RESULT 12
LIVJ_ECOLI STANDARD; PRT; 367 AA.
ID LIVJ_ECOLI
AC P02917; P76698;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE LEU/ILE/VAL-BINDING PROTEIN PRECURSOR (LIV-BP).
GN LIVJ.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 85234531.
RA Landick R., Oxender D.L.;
RT "The complete nucleotide sequences of the Escherichia coli Liv-BP and
RT LS-BP genes. Implications for the mechanism of high-affinity
RT branched-chain amino acid transport";
RT J. Biol. Chem. 260:8257-8261(1985).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE: 86086153.
RA Antonucci T.K., Landick R., Oxender D.L.;
RT "The leucine binding proteins of Escherichia coli as models for
RT studying the relationships between protein structure and function";
RT J. Cell. Biochem. 29:209-216(1985).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90307651.
RA Adams M.D., Wagner L.M., Graddis T.J., Landick R., Antonucci T.K.,
RA Gibson A.L., Oxender D.L.;
RT "Nucleotide sequence and genetic characterization reveal six
RT essential genes for the Liv-I and LS transport systems of Escherichia
RT coli";
RT J. Biol. Chem. 265:11436-11443(1990).
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN-K12 / MG1655;
RX MEDLINE: 94316500.
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes";
RT Nucleic Acids Res. 22:2576-2586(1994).
RN [5]
RP SEQUENCE OF 24-367.
RX STRAIN-K12;
RX MEDLINE: 77225636.
RA Ovchinnikov Y.A., Aldanova N.A., Grinkevich V.A., Arzamazova N.M.,
RA Moroz I.N.;
RT "The primary structure of a leu, ile and val (liv)-binding protein
RT from Escherichia coli";
RL FEBS Lett. 78:313-316(1977).
RN [6]
RP SEQUENCE OF 24-35.
RX STRAIN-K12 / EMG2;
RX MEDLINE: 97443975.

RA Link A.J., Robison K., Church G.M.;
 RT Comparing the predicted and observed properties of proteins encoded
 RT in the genome of *Escherichia coli* K-12.";
 RL Electrophoresis 18:1259-1313(1997).
 RN 171
 RC SEQUENCE OF 24-35.
 RA STRAIN-K12 / W3110;
 RA Fritiger S., Hughes G.J., Pasquani C., Hochstrasser D.F.;
 RL Submitted (FEB-1996) to the SWISS-PROT data bank.
 RN 181
 RP X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS).
 RX MEDLINE: 89199638.
 RA Beck J.S., Saper M.A., Quijcho F.A.;
 RT Periplasmic binding protein structure and function. Refined X-ray
 RT structures of the leucine/isoleucine/valine-binding protein and its
 RT complex with leucine.";
 RL J Mol. Biol. 206:171-191(1989).
 CC - FUNCTION: THIS PROTEIN IS A COMPONENT OF THE LEUCINE, ISOLEUCINE,
 CC VALINE, (THREONINE) TRANSPORT SYSTEM, WHICH IS ONE OF THE TWO
 CC PERIPLASMIC BINDING PROTEIN-DEPENDENT TRANSPORT SYSTEMS OF THE
 CC HIGH-AFFINITY TRANSPORT OF THE BRANCHED-CHAIN AMINO ACIDS.
 CC - SUBCELLULAR LOCATION: PERIPLASMIC.
 CC - SIMILARITY: BELONGS TO THE LEUCINE-BINDING PROTEIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC 1
 DR EMBL: J05516; AAA83881.1; -;
 DR EMBL: M29377; AAA24075.1; -;
 DR EMBL: U00039; CAB34652.1; ALT_INIT.
 DR EMBL: AE000422; AAC76485.1; ALT_INIT.
 DR EMBL: A03415; BLEC.
 DR PIR: A23576; A23576.
 DR PIR: A37074; A37074.
 DR RDB: 2LIV; 12-JUL-89.
 DR SWISS-2DPAGE: P02917; COLI.
 DR ECG2DBASE: D040.7; 6TH EDITION.
 DR ECGENE: EGI0539; LIVU.
 DR INTERPRO: IPR000709; -;
 DR INTERPRO: IPR001828; -;
 DR PIRAM: PF01094; ANF_receptor; 1.
 DR PRINTS: PR00337; LEU1EVALBP.
 KM Antic-acid transport: Transport; Periplasmic; Signal: 3D-structure.
 FT SIGNL 1 23
 FT CHAIN 24 367
 FT DI_SPLID 76 101
 FT CONFLICT 3 3
 FT CONFLICT 12 12
 FT CONFLICT 16 16
 FT CONFLICT 70 70
 FT STRAND 26 32
 FT HELIX 39 58
 FT TURN 59 61
 FT STRAND 62 62
 FT TURN 63 64
 FT STRAND 65 65
 FT STRAND 67 73
 FT TURN 75 76
 FT HELIX 78 90
 FT TURN 91 92
 FT STRAND 95 98
 FT HELIX 102 114
 FT TURN 115 116
 FT STRAND 118 121
 FT HELIX 127 129
 FT STRAND 137 139
 FT HELIX 144 157
 FT TURN 158 159

FT STRAND 164 168
 FT HELIX 172 186
 FT TURN 187 189
 FT STRAND 192 197
 FT TURN 200 201
 FT HELIX 206 214
 FT TURN 215 216
 FT STRAND 219 223
 FT HELIX 226 238
 FT TURN 239 240
 FT STRAND 244 248
 FT HELIX 244 251
 FT HELIX 254 264
 FT TURN 265 266
 FT STRAND 268 271
 FT HELIX 275 277
 FT HELIX 279 288
 FT TURN 289 292
 FT TURN 295 296
 FT HELIX 299 314
 FT TURN 315 316
 FT HELIX 320 329
 FT STRAND 332 334
 FT TURN 335 336
 FT STRAND 337 339
 FT STRAND 341 341
 FT TURN 343 344
 FT STRAND 347 347
 FT STRAND 353 357
 FT TURN 359 360
 FT STRAND 363 365
 SQ SEQUENCE 367 AA; 39046 MW; 4AF17D3C3AA74D9C CRC64;

Query Match 87.1%; Score 27; DB 1; Length 367;
 Best Local Similarity 83.3%; Pred. No. 44;
 Matches 5; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 1 MNYGK 6
 Db 1 MNTKG 6

RESULT 13
 SABC_BACSU STANDARD; PRT; 473 AA.
 ID SABC_BACSU
 AC P05655; P70984;
 DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DT 01-NOV-1988 (Rel. 09, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE LEVANSUCRASE PRECURSOR (EC 2.4.1.10) (BETA-D-FRUCTOFURANOSYL
 DE TRANSFERASE) (SUCROSE 6-FRUCTOSYL TRANSFERASE).
 GN SABC.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RX MEDLINE: 85295507.
 RA Steimetz M., Le Cog D., Aymerich S., Gonzy-Treboul G., Gay P.;
 RT "The DNA sequence of the gene for the secreted Bacillus subtilis
 RT enzyme levansucrase and its genetic control sites.";
 RL Mol. Gen. Genet. 200:220-228(1985).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Denizot F.C.;
 RN Submitted (APR-1997) to the EMBL/GenBank/DBJ databases.
 RC [3]
 RP STRAIN-168 / PY79;
 RX MEDLINE: 84178454.
 RA Fouet A., Arnaud M., Klier A., Rapoport G.;

RT "Characterization of the precursor form of the exocellular
RT levansucrase from *Bacillus subtilis*."
RT Biochem. Biophys. Res. Commun. 119:795-800(1984).
RN (4)
RX SEQUENCE OF 1-68 FROM N.A.
RP MEDLINE: 8708406.
RA Shimotsu H., Henner D.J.:
RT "Modulation of *Bacillus subtilis* levansucrase gene expression by
RT sucrose and regulation of the steady-state mRNA level by *sacU* and
RT *sacQ* genes."
RL J. Bacteriol. 168:380-388(1986).
CC -1- CATALYTIC ACTIVITY: SUCROSE + (2,6-BETA-D-FRUCTOSYL)(N) -
CC GLUCOSE + (2,6-BETA-D-FRUCTOSYL)(N+1) (OTHER SUGARS CAN
CC ALSO ACT AS FRUCTOSYL ACCEPTORS).
CC -1- SUBCELLULAR LOCATION: SECRETED.
CC -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M14202; AAA22725.1; -
DR EMBL: 294043; CAB08015.1; -
DR EMBL: K01987; AAA22724.1; -
DR EMBL: X02730; CAA26513.1; -
DR EMBL: 299121; CAB15450.1; -
DR PIR: A25040; A25040.
DR PIR: S07309; S07309.
DR SUBTILIST: BG10388; SABC.
DR Transferase: Glycosyltransferase; Signal.
KW SIGNAL
FT SIGNAL 1 29
FT CHAIN 30 473 LEVANSUCRASE.
FT CONFLICT 12 12 V -> I (IN REF. 3).
FT SEQUENCE 473 AA; 52971 MW; 3FBF2F571B4D5B0 CRC64;
SQ
Query Match 87.1%; Score 27; DB 1; Length 473;
Best Local Similarity 66.7%; Pred. No. 56;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNVKGK 6
DB 449 LNIKGR 454
RESULT 14
ID CVA2_HUMAN STANDARD; PRT; 476 AA.
AC 008462;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE ADENYLATE CYCLASE, TYPE II (EC 4.6.1.1) (ATP PYROPHOSPHATE-LYASE)
DE (ADENYLYL CYCLASE) (FRAGMENT).
GN ADCY2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE-BRAIN;
RX MEDLINE: 93052215.
RA Stengel D., Parma J., Gannage M.-H., Roedel N., Matzel M.-G.,
RA Barokki R., Hanoune J.:
RT "Different chromosomal localization of two adenylyl cyclase genes
RT expressed in human brain."
RL Hum. Genet. 90:126-130(1992).
CC -1- FUNCTION: THIS A MEMBRANE-BOUND, CALMODULIN-INSENSITIVE ADENYLYL
CC CYCLASE.

CC -1- CATALYTIC ACTIVITY: ATP - 3',5'-CYCLIC AMP + PYROPHOSPHATE.
CC -1- ENZYME REGULATION: INSENSITIVE TO CA(2+)/CALMODULIN. STIMULATED BY
CC THE G PROTEIN BETA AND GAMMA SUBUNIT COMPLEX (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN.
CC -1- TISSUE SPECIFICITY: EXPRESSED IN BRAIN.
CC -1- DOMAIN: COMPOSED OF TWO HOMOLOGOUS DOMAINS.
CC -1- SIMILARITY: BELONGS TO ADENYLYL CYCLASE CLASS-4/GUANYLYL CYCLASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X74210; CAA52282.1; -
DR EMBL: L21993; AAA64923.1; -
DR HSSP: P26769; IAB8.
DR MIM: 103071; -
DR INTERPRO: IPR001054; -
DR PFW: P00211; guanylate-cyc: 1.
DR PROSITE: P500452; GUANYLATE_CYCLASES; 1.
KW Lyase; CAMP synthesis; Transmembrane; Glycoprotein; Duplication.
FT NON_TER 1 1
FT TRANSMEM 13 7 POTENTIAL.
FT TRANSMEM 65 37 POTENTIAL.
FT TRANSMEM 119 140 POTENTIAL.
FT TRANSMEM 148 165 POTENTIAL.
FT TRANSMEM 186 206 POTENTIAL.
FT DOMAIN 207 476 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 98 98 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 101 101 N-LINKED (GLCNAC...) (POTENTIAL).
FT SEQUENCE 476 AA; 53430 MW; 33C6019B10892F5E CRC64;
SQ
Query Match 87.1%; Score 27; DB 1; Length 476;
Best Local Similarity 83.3%; Pred. No. 57;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 1 MNVKGK 6
DB 448 INVKGK 453
RESULT 15
ID RT04_PROWI STANDARD; PRT; 511 AA.
AC P46743;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE MITOCHONDRIAL RIBOSOMAL PROTEIN S4.
GN RPS4.
OS Prototheca wickerhamii.
OC Mitochondrion.
OC Eukaryota; Viridiplantae; Chlorophyta; Trebouxiophyceae; Chlorellales;
OC Chlorellaceae; Prototheca.
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN-263-11;
RX MEDLINE: 94180393.
RA Wolff G., Plante I., Lang B.F., Kueck U., Burger G.:
RT "Complete sequence of the mitochondrial DNA of the chlorophyte *algae*
RT *Prototheca wickerhamii*. Gene content and genome organization."
RL J. Mol. Biol. 237:75-86(1994).
CC -1- SUBCELLULAR LOCATION: MITOCHONDRIAL.
CC -1- SIMILARITY: BELONGS TO THE S4P FAMILY OF RIBOSOMAL PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC -----
DR ENRL: U02970; AAD12659.1; -
DR MENDEL; 12232; FROM: IPR54; 1.
DR INTERPRO: IPR001912; -
DR INTERPRO: IPR002942; -
DR PAM; PF01479; S4; 1.
DR PROSITE: PS00632; RIBOSOMAL_S4; FALSE_NEG.
KW Mitochondrial protein; Mitochondrion
SQ SEQUENCE 511 AA; 59213 MW; 6DAA46528CB09576 CRC64;

Query Match 87.1%; Score 27; DB 1; Length 511;
Best Local Similarity 66.7%; Pred. No. 61;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNVKGK 6
: : : : :
Db 440 LNIKKG 445

Search completed: December 14, 2000, 10:14:58
Job time: 223 sec

FT	MUTAGEN	1	46	MISSING:	20% INCREASE IN ACTIVITY.
FT	MUTAGEN	1	89	MISSING:	20% INCREASE IN ACTIVITY.
FT	MUTAGEN	1	90	MISSING:	50% DECREASE IN ACTIVITY.
FT	MUTAGEN	1	91	MISSING:	75% DECREASE IN ACTIVITY.
FT	MUTAGEN	1	92	MISSING:	82% DECREASE IN ACTIVITY.
FT	MUTAGEN	1	93	MISSING:	15% DECREASE IN ACTIVITY.
FT	MUTAGEN	1	96	MISSING:	NO ACTIVITY.
FT	MUTAGEN	1	101	MISSING:	NO ACTIVITY.
FT	MUTAGEN	318	318	I->:	100% DECREASE IN ACTIVITY.
FT	MUTAGEN	374	376	MISSING:	>95% DECREASE IN ACTIVITY.
FT	MUTAGEN	375	375	MISSING:	70% DECREASE IN ACTIVITY.
FT	MUTAGEN	375	375	N->:	NO CHANGE IN ACTIVITY.
FT	MUTAGEN	376	376	V->:	NO CHANGE IN ACTIVITY.
SO	SEQUENCE	376 AA;	444/71 MM;	13C490CDBF2A2ADF	CRC6d;

Query Match	100.0%	Score 31;	DB 6;	Length 376;
Best Local Similarity	100.0%	Pred. No. 30;		
Matches	6;	Conservative	0;	Mismatches 0;
			Indels	0;
			Gaps	0;

Qy	1	MNVKGK	6
Db	1	MNVKGK	6

RESULT	2
Q92371	
ID	Q92371
PRELIMINARY;	
PRT;	213 AA

DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE REGULATIVE TRANSCRIPTIONAL REGULATOR.

05 *Helicobacter pylori* J99 (*Campylobacter pylori* J99).
0C Bacteria; Proteobacteria; epsilon subdivision; *Helicobacter* group;
0C *Helicobacter*.

RP SEQUENCE FROM N.A.
RX MEDLINE; 99120557.
RA Alm R.A., Ling L.-S.L., Molr D.T., King B.L., Brown E.D., Doig P.C.,
RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carnel G.,
RA Timmimo P.J., Caruso A., Uria-Nikelsen M., Mills D.M., Ives C.,
RA Gibson R., Werberg D., Mills S.D., Jiang Q., Taylor D.E., Voyls G.F.,
RA Target T.V.,
RT Genomic sequence comparison of two unrelated isolates of the human
RT gastric pathogen *Helicobacter pylori*.";

RL Nature397:176-180(1999).
 DR EMBL: AE001567; AAD07036.1; -.
 DR EMBL: AE001551; AAD06868.1; -.
 DR HSSP: P06628; 2FSP.
 DR INTERPRO: IPR001789; -.
 DR INTERPRO: IPR001867; -.
 DR Pfam: PF00072; response_reg.1.
 DR Pfam: PF00486; trans_reg.4.1.
 DR Qc SEQUENCE 213 AA: 2482 MW: C32BFLADBD323347C CXC66;

Query Match	90.3%	Score 28;	DB 2;	Length 213;	
Best Local Similarity	83.3%	Pred. No. 82;			
Matches	5;	Conservative	1;	Mismatches	0;
				Indels	0;
				Gaps	0;

QY	1	MNVK GK	6
		: 11111	
Db	137	LN VK GK	142

RESULT	3
Q9TLM8	
ID	Q9TLM8
PRELIMINARY;	
PRT;	266 AA

DT	01-MAY-2000 (TREMBLrel. 13, Created)
DT	01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE HYPOTHEETICAL 29.3 KDA PROTEIN.
GN TRPA.
OS Cyanidium caldarium (Galdieria sulphuraria).
OC Chloroplast.
CC Eukaryote; Rhodophyta; Bangiophyceae; Porphyridiales; Porphyridiaceae;
OC Cyanidium.
RN [1]
RP SEQUENCE FROM N.A.
RA Gloeckner G., Rosenthal A., Valentin K.;
RC STRAIN-RK1
RT "Reconstruction of the ancient red algal plastid genome: Structure,
RL submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF022186; AAF12944.1; -.
DR INTERPRO: IPR002028; -.
DR INTERPRO: IPR003009; -.
DR PFAM: PF00290; trp-syntA; 1.
DR PROSITE: PS00167; TRP SYNTHASE ALPHA; 1.
KW Hypothetical protein; Chloroplast.
SO SEQUENCE 266 AA; 29344 MW; E5667FDD3007168D CRC64;

Query Match	90.3%	Score 28;	DB 8;	Length 266;
Best Local Similarity	83.3%;	Pred. No. 1e+02;		
Matches	5;	Conservative	1;	Mismatches 0;
			Indels	0;
			Gaps	0;

QY	1	MNVK GK	6
		:	
Db	84	MNLK GK	89

RESULT	4	
059163		
ID	059163	PRELIMINARY; PRT; 369 AA

DT 01-AUG-1998 (TREMblrel. 07, Created)
DT 01-AUG-1998 (TREMblrel. 07, Last sequence update)
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE HYPOTHEtical 42.3 kDa protein PH194.

OS *Pyrococcus horikoshii*,
OC Archaea; Euryarchaeota; Thermococcales; Pyrococcus
RN [4]
RP SEQUENCE FROM N.A.

RC STRAIN-OT3;
RX MEDLINE: 98344137.
RA Kawarabayashi Y., Sawada M., Horikawa H., Hinkawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohnuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushioka N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.,
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, *Pyrococcus horikoshii* OT3.",
RL DNA Res. 5:55-76(1998).
DR EMBL: AP000006; BAA3601.1; -
KW Hypothetical protein.
QW SEQUENCE 369 AA: 42349 MW: BD79BEC7889F1388 CRC64;

Query Match	90.3%	Score 28;	DB 1;	Length 369;
Best Local Similarity	83.3%	Pred. NO. 1.4e+02;		
Matches	5;	Conservative	1;	Mismatches 0;
			Indels	0;
			Gaps	0;

```
QY      1 MNVKGK 6
          :|||||
Db      216 LNVKGK 221
```

RESULT	5	
Q9YCR2		
ID	Q9YCR2	PRELIMINARY;
		PRT; 458 AA

AC Q9YCR2;
 DT 01-NOV-1999 (TRENBLREL. 12, Created)
 DT 01-NOV-1999 (TRENBLREL. 12, Last sequence update)
 DT 01-MAY-2000 (TRENBLREL. 13, Last annotation update)
 DE HYPOTHETICAL 52.0 KDA PROTEIN APE1199.
 GN APE1199.
 OS Aeropyrum pernix.
 OC Archaea; Crenarchaeota; Desulfurococcales; Desulfurococcaceae;
 OC Aeropyrum.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K1;
 RX MEDLINE: 99310339.
 RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Halkawa Y.,
 Jin-no K., Takahashi M., Sekine M., Baba S., Anka A., Kosugi H.,
 Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
 Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
 Yamazaki J., Kushiida N., Oguchi A., Aoki K., Kubota K., Nakamura Y.,
 Nomura N., Sako Y., Kikuchi H.;
 RA "Complete genome sequence of an aerobic hyper-thermophilic
 RT crenarchaeon, Aeropyrum pernix K1.";
 RL DNA Res. 6:83-101(1999).
 DR EMBL: AP000061; BAA80185.1; -.
 KW Hypothetical protein.
 SQ SEQUENCE 458 AA; 51974 MW; 5795FDA58016304 CRC64;

Query Match 90.3%; Score 28; DB 1; Length 458;
 Best Local Similarity 83.3%; Pred. No. 1.8e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKGK 6
 :|||||
 DB 31 LNVKGR 36

RESULT 6
 ID 085325 PRELIMINARY: PRT; 463 AA.
 AC 085325;
 DT 01-NOV-1998 (TRENBLREL. 08, Created)
 DT 01-NOV-1998 (TRENBLREL. 08, Last sequence update)
 DT 01-JUN-2000 (TRENBLREL. 14, Last annotation update)
 DE HYPOTHETICAL 49.2 KDA PROTEIN.
 GN SPI4_P.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-LT2;
 RX MEDLINE: 96296059.
 RA Wong K.K., McClelland M., Stillwell L.C., Sisk E.C., Thurston S.J.,
 Saffer J.D.;
 RA "Identification and sequence analysis of a 27-kilobase chromosomal
 RT fragment containing a Salmonella pathogenicity island located at 92
 RT minutes on the chromosome map of Salmonella enterica serovar
 RT typhimurium LT2.";
 RL Infect. Immun. 66:3365-3371(1998).
 DR EMBL: AF060869; AAC26653.1; -.
 DR INTERPRO: IPR000150; -.
 DR PROSITE: PS01228; COF_1; UNKNOWN_1.
 KW Hypothetical protein.
 SQ SEQUENCE 463 AA; 49218 MW; D62AD07F6375E428 CRC64;

Query Match 90.3%; Score 28; DB 2; Length 463;
 Best Local Similarity 83.3%; Pred. No. 1.8e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKGK 6
 :|||||
 DB 417 MNVGRK 422

RESULT 7
 ID 035812 PRELIMINARY: PRT; 749 AA.
 AC 035812;
 DT 01-JAN-1998 (TRENBLREL. 05, Created)
 DT 01-JAN-1998 (TRENBLREL. 05, Last sequence update)
 DT 01-JUN-2000 (TRENBLREL. 14, Last annotation update)
 DE PEX PROTEIN.
 GN PEX.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-INCISOR.
 RX MEDLINE: 97343325.
 RA Francis F., Strom T.M., Hennig S., Boeddrich A., Lorenz B.,
 Brandau O., Mohnlke K.L., Cagnoli M., Steffens C., Klages S.,
 Borzym K., Pohl T., Oudet C., Econs M.J., Rowe P.S., Reinhardt R.,
 Meltinger T., Lebrach H.;
 RA "Genomic organization of the human PEX gene mutated in X-linked
 RT dominant hypophosphatemic rickets.";
 RL Genome Res. 7:573-585(1997).
 DR EMBL: AJ001637; CAA04890.1; -.
 DR INTERPRO: IPR000130; -.
 DR INTERPRO: IPR000718; -.
 DR PFM: PF01431; Peptidase_M13; 1.
 DR PRINTS: PR00786; NEPRILysin.
 DR PROSITE: PS00142; ZINC_PROTEASE; UNKNOWN_1.
 SQ SEQUENCE 749 AA; 86305 MW; 690B8E82609C1A63 CRC64;

Query Match 90.3%; Score 28; DB 1; Length 749;
 Best Local Similarity 83.3%; Pred. No. 2.8e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKGK 6
 :|||||
 DB 632 LNVKGR 637

RESULT 8
 ID 09VZP5 PRELIMINARY: PRT; 932 AA.
 AC 09VZP5;
 DT 01-MAY-2000 (TRENBLREL. 13, Created)
 DT 01-MAY-2000 (TRENBLREL. 13, Last sequence update)
 DT 01-JUN-2000 (TRENBLREL. 14, Last annotation update)
 DE CG10840 PROTEIN.
 GN CG10840.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY;
 RX MEDLINE: 20196006.
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 Brandon R.C., Rogers Y.-H.C., Blazei R.G., Champe M., Pfeiffer B.D.,
 Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,
 Ballif J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotlier P.,
 Burks K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,
 Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Gloder A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
 RA Jatalai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "the genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003478; AAF4774.1; -;
 DR FLYBASE: FBgn0035447; CG10840.
 DR INTERPRO: IPR000795; -;
 DR PRAM: PF00009; GTP_EFTU; 1;
 DR PRINTS: PRO0315; ELONGACTNCT.
 SQ SEQUENCE 932 AA; 102828 MW; 3899C68FF38ED23F CRC64;

Query Match 90.3%; Score 28; DB 5; Length 932;
 Best Local Similarity 83.3%; Pred. No. 3.5e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKG 6
 Db 62 LNVKG 67

RESULT 9
 ID 09VNS7 PRELIMINARY; PRT; 967 AA.
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
 DE G6L7387 PROTEIN.
 GN G6L7387.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota: Metazoa: Arthropoda: Tracheata: Hexapoda: Insecta:
 OC Eterygota: Neoptera: Endopterygota: Diptera: Brachycera: Muscomorpha:
 OC Ephydroidea: Drosophilidae: *Drosophila*.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BERKELEY.
 RX MEDLINE: 20196006.
 RA Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amandalides P.G., Scherter S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandal M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazet R.G., Champe M., Pfeiffer B.D.,
 RA Wen K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mikos G.L.G.,
 RA Aroll J.F., Agbayani A., An H.-U., Andrews-Plannkoch C., Baldwin D.,
 RA Bailew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Bogkova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos J., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,

RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Gloder A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,
 RA Jatalai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "the genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 DR EMBL: AE003605; AAF52093.1; -;
 DR FLYBASE: FBgn0037276; CG17387.
 SQ SEQUENCE 967 AA; 112254 MW; 109A91FF6703E941 CRC64;

Query Match 90.3%; Score 28; DB 5; Length 967;
 Best Local Similarity 83.3%; Pred. No. 3.7e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKG 6
 Db 917 LNVKG 922

RESULT 10
 ID 09QX12 PRELIMINARY; PRT; 1135 AA.
 AC 09QX12;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
 DE GTPASE ACTIVATING PROTEIN SYNGAP-C.
 OS *Rattus norvegicus* (Rat).
 OC Eukaryota: Metazoa: Chordata: Vertebrata: Euteleostomi:
 OC Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Murinae: *Rattus*.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=HIPPOCAMPUS;
 RA Kim J.H., Liao D., Lau L.-F., Huganir R.L.;
 RT "SynGAP: a Synaptic RasGAP that Associates with the PSD-95/SAP90
 Protein Family.";
 RL Neuron 20:683-691(1998).
 DR EMBL: AF050183; AAC40082.1; -;
 DR INTERPRO: IPR001849; -;
 DR INTERPRO: IPR001936; -;
 DR PRAM: PF00169; PH; 1.
 DR PRAM: PF00616; RasGAP; 1.
 DR POSITIVE: PSS0003; PH_DOMAIN; 1.
 DR POSITIVE: PSS0018; RAS-GTPASE_ACTIV_2; 1.
 SQ SEQUENCE 1135 AA; 124718 MW; 8BA760EEDE45E9DA CRC64;

Query Match 90.3%; Score 28; DB 11; Length 1135;
 Best Local Similarity 83.3%; Pred. No. 4.3e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MNVKG 6
 Db 266 LNVKG 271

```

RESULT 11
090M12 PRELIMINARY: PRT: 1166 AA.
AC 090M12:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE SYNGAP-B1.
GN SYNGAP-B1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RA Suzuki T.;
RT "SYNGAP-B1."
RL Submitted (AUG-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB016962; CAB19493.1; -.
DR INTERPRO: IPR001849; -.
DR INTERPRO: IPR001936; -.
DR PFAM: PF00169; PH; 1.
DR PFAM: PF00616; RasGAP; 1.
DR PROSITE: PS50003; PH_DOMAIN; 1.
DR PROSITE: PS50018; RAS_GTPASE_ACTIV_2; 1.
SO SEQUENCE 1166 AA; 127744 MW; 5804EE4163AE4950 CRC64;

```

```

Query Match          90.3%; Score 28; DB 11; Length 1166;
Best Local Similarity 83.3%; Pred. No. 4.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MNVKGK 6
Db 322 LNVKKG 327

```

```

RESULT 12
090G2 PRELIMINARY: PRT: 1179 AA.
AC 090G2:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE DJ570F3.1 (HOMOLOG OF THE RAT SYNAPTIC RAS GTPASE-ACTIVATING PROTEIN
DJ570F3.1.
GN DJ570F3.1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
RN [1]
RP SEQUENCE FROM N.A.
RA Mashreghi-Mohammadi M.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL050332; CAB63776.1; -.
DR INTERPRO: IPR001849; -.
DR INTERPRO: IPR001936; -.
DR PFAM: PF00169; PH; 1.
DR PFAM: PF00616; RasGAP; 1.
DR PROSITE: PS50003; PH_DOMAIN; 1.
DR PROSITE: PS50018; RAS_GTPASE_ACTIV_2; 1.
SO NON-TER 1179 AA; 129242 MW; E7C3FE44DAC599A0 CRC64;

```

```

Query Match          90.3%; Score 28; DB 4; Length 1179;
Best Local Similarity 83.3%; Pred. No. 4.4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MNVKGK 6
Db 424 LNVKKG 429

```

```

RESULT 13
090X02 PRELIMINARY: PRT: 1249 AA.
AC 090X02:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE SYNGAP-B.
GN Rattus norvegicus (Rat).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-HIPPOCAMPUS;
RA Kim J.H., Liao D., Lau L.-F., Huganir R.L.;
RT "SYNGAP: a synaptic RasGAP that Associates with the PSD-95/SAP90
Protein Family."
RL Neuron 20:683-691(1998).
DR EMBL: AF058790; AAC63511.1; -.
DR INTERPRO: IPR001849; -.
DR INTERPRO: IPR001936; -.
DR PFAM: PF00169; PH; 1.
DR PFAM: PF00616; RasGAP; 1.
DR PROSITE: PS50003; PH_DOMAIN; 1.
DR PROSITE: PS50018; RAS_GTPASE_ACTIV_2; 1.
SO SEQUENCE 1249 AA; 137357 MW; DC549FCF1624BD8A CRC64;

```

```

Query Match          90.3%; Score 28; DB 11; Length 1249;
Best Local Similarity 83.3%; Pred. No. 4.7e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MNVKGK 6
Db 380 LNVKKG 385

```

```

RESULT 14
090UH6 PRELIMINARY: PRT: 1293 AA.
AC 090UH6:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE SYNGAP-A.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-HIPPOCAMPUS;
RA Kim J.H., Liao D., Lau L.-F., Huganir R.L.;
RT "SYNGAP: a synaptic RasGAP that Associates with the PSD-95/SAP90
Protein Family."
RL Neuron 20:683-691(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY;
RA Chen H.-J., Kennedy M.B.;
RT "Identification and cloning of a novel 130 kd protein containing a ras
GTPase-activating domain from the rat forebrain postsynaptic
RT density."
RT Abstr. Soc. Neurosci. 1466:1-1(0).
DR EMBL: AF058789; AAC63510.1; -.
DR EMBL: AF048976; AAC08071.1; -.
DR INTERPRO: IPR001849; -.
DR INTERPRO: IPR001936; -.
DR PFAM: PF00169; PH; 1.
DR PFAM: PF00616; RasGAP; 1.
DR PROSITE: PS50003; PH_DOMAIN; 1.
DR PROSITE: PS50018; RAS_GTPASE_ACTIV_2; 1.

```

SQ SEQUENCE 1293 AA; 143113 MW; CEA63FA7E561413 CRC64;

Query Match 90.3%; Score 28; DB 11; Length 1293;

Best Local Similarity 83.3%; Pred. No. 4.8e+02; Mismatches 0; Indels 0; Gaps 0;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNVKGK 6

Db 424 LNVKGK 429

RESULT 15

ID 090947 PRELIMINARY; PRT; 198 AA.

AC 090947; 01-MAY-2000 (T-EMBLrel. 13, Created)

DT 01-MAY-2000 (T-EMBLrel. 13, Last sequence update)

DT 01-JUN-2000 (T-EMBLrel. 14, Last annotation update)

DE GP68 (FRAGMENT).

GN MSPC.

OS Leishmania infantum.

OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Leishmania.

RN 111

RP SEQUENCE FROM N.A.

RC STRAIN-MCAN/FR/82/PHAROAH;

RA Mauricio I.L., Howard M.R., Stothard J.R., Miles M.A.;

RT "Genomic diversity in the Leishmania donovani complex";

RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.

DR EMBL: AJ010248; CAB51797.1; -.

DR INTERPRO: IPR001577; -.

DR PRINTS: PR00782; LSHMANOLYSIN.

FT NON-TER 1 1

FT NON-TER 198 198

SO SEQUENCE 198 AA; 21856 MW; 63468A7C174972EC CRC64;

Query Match 87.1%; Score 27; DB 5; Length 198;

Best Local Similarity 66.7%; Pred. No. 1.3e+02; Mismatches 0; Indels 0; Gaps 0;

QY 1 MNVKGK 6

Db 11 MNIRGK 16

Search completed: December 14, 2000, 10:12:27
Job time: 121 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 14, 2000, 10:11:10 ; Search time 36.07 seconds
(without alignments)
5.688 Million cell updates/sec

Title: US-09-051-034A-13

Perfect score: 29
Sequence: 1 MVMKKG 6

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues

Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A_Geneseq_36.*
1: /SIDSI/gcgcdata/geneseq/geneseqp/AA1980.DAT.*
2: /SIDSI/gcgcdata/geneseq/geneseqp/AA1981.DAT.*
3: /SIDSI/gcgcdata/geneseq/geneseqp/AA1982.DAT.*
4: /SIDSI/gcgcdata/geneseq/geneseqp/AA1983.DAT.*
5: /SIDSI/gcgcdata/geneseq/geneseqp/AA1984.DAT.*
6: /SIDSI/gcgcdata/geneseq/geneseqp/AA1985.DAT.*
7: /SIDSI/gcgcdata/geneseq/geneseqp/AA1986.DAT.*
8: /SIDSI/gcgcdata/geneseq/geneseqp/AA1987.DAT.*
9: /SIDSI/gcgcdata/geneseq/geneseqp/AA1988.DAT.*
10: /SIDSI/gcgcdata/geneseq/geneseqp/AA1989.DAT.*
11: /SIDSI/gcgcdata/geneseq/geneseqp/AA1990.DAT.*
12: /SIDSI/gcgcdata/geneseq/geneseqp/AA1991.DAT.*
13: /SIDSI/gcgcdata/geneseq/geneseqp/AA1992.DAT.*
14: /SIDSI/gcgcdata/geneseq/geneseqp/AA1993.DAT.*
15: /SIDSI/gcgcdata/geneseq/geneseqp/AA1994.DAT.*
16: /SIDSI/gcgcdata/geneseq/geneseqp/AA1995.DAT.*
17: /SIDSI/gcgcdata/geneseq/geneseqp/AA1996.DAT.*
18: /SIDSI/gcgcdata/geneseq/geneseqp/AA1997.DAT.*
19: /SIDSI/gcgcdata/geneseq/geneseqp/AA1998.DAT.*
20: /SIDSI/gcgcdata/geneseq/geneseqp/AA1999.DAT.*
21: /SIDSI/gcgcdata/geneseq/geneseqp/AA2000.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	28	96.6	185	17	W03548
2	27	93.1	415	20	W97389
3	27	93.1	415	21	Y36505
4	27	93.1	427	20	W97388
5	27	93.1	427	21	Y56504
6	26	89.7	64	16	R78382
7	26	89.7	187	21	Y82505
8	26	89.7	254	15	R50610
9	26	89.7	264	16	R67735
10	26	89.7	268	16	R78388
11	26	89.7	268	19	W71250
12	26	89.7	339	21	Y68661

13	26	89.7	341	21	Y68664	Amino acid sequenc
14	26	89.7	345	21	Y68668	Amino acid sequenc
15	26	89.7	343	21	Y68665	Amino acid sequenc
16	26	89.7	390	21	Y68648	Amino acid sequenc
17	26	89.7	391	21	Y68650	Amino acid sequenc
18	26	89.7	392	21	Y68654	Amino acid sequenc
19	26	89.7	394	21	Y68659	Amino acid sequenc
20	26	89.7	472	15	R56447	TWV replicon-encod
21	26	89.7	472	19	W37940	Amino acid sequenc
22	26	89.7	472	21	Y68328	Amvotrophic lateta
23	26	89.7	485	19	W71249	protein encoded by
24	26	89.7	1049	18	W22159	ApvIIIIB protein.
25	26	89.7	1049	21	Y51412	A. pleuropneumonia
26	26	89.7	1244	15	R54781	Leukotoxin AppIIA
27	25	86.2	128	20	Y29977	Human electron tra
28	25	86.2	146	20	Y59740	Human normal ovari
29	25	86.2	336	15	R59739	Pseudomonas glutam
30	25	86.2	347	12	R11069	1203 antigen seque
31	25	86.2	392	21	Y68658	Amino acid sequenc
32	25	86.2	483	11	R09253	c-PA variant dl-44
33	25	86.2	483	16	R70883	Human tissue PA va
34	25	86.2	483	16	R70890	Human tissue PA va
35	25	86.2	483	16	R70849	Human tissue PA va
36	25	86.2	483	16	R70855	Human tissue PA va
37	25	86.2	894	20	W68810	WPA protein. Bac
38	25	86.2	1634	21	Y52023	M. jannaschli MJO8
39	25	86.2	1634	21	Y51652	M. jannaschli MJO8
40	25	86.2	1670	17	R97048	DNA polymerase, co
41	25	86.2	1670	17	R94616	DNA polymerase, P
42	25	86.2	1670	17	R85599	DNA polymerase, P
43	25	86.2	1670	19	W20049	KODI thermostable
44	25	86.2	1670	18	W41314	Full length heat-r
45	24	82.8	9	20	Y36561	Fragment of human

ALIGNMENTS

RESULT 1	
ID W03548	standard; Protein: 185 AA.
AC W03548;	
DT 21-APR-1997	(first entry)
DE	ORF-2 protein sequence from BamHI fragment of HVT.
XX	
XX	Herpes virus of turkey; open reading frame; ORF; homology; vector;
KW	avian herpes virus; recombinant viral vaccine; intergenic region; IBDV;
KW	cytomegalovirus immediate early promoter; US5 gene; repeat region; ILTV;
KW	antigen; infectious bursal disease virus; Marek's disease virus; MDV;
KW	infectious laryngotracheitis virus; avian anaemia virus; vaccination;
KW	infectious bronchitis virus; IBV; poultry; Gumboro disease;
XX	Newcastle disease.
XX	
OS	Herpesvirus of turkey.
XX	
PN	EP719864-A2.
XX	
PD	03-JUL-1996.
XX	
PF	28-DEC-1995; 95EP-0402970.
XX	
PR	30-DEC-1994; 94FR-0016017.
XX	
PA	(INMR) RHONE MERIEUX SA.
XX	
PI	Audonnet JCF, Budiot MJM, Dartell RJ, Duinat CV;
XX	Laplace ELP, Riviere MAE;
XX	
DR	WPI: 1996-364150/37.
XX	
DR	N-PSDB: T39309.

XX Live recombinant avian vaccine - comprises herpes virus as vector
PT and having sequence encoding antigenic polypeptide inserted between
PT U155 gene and repeat region
XX
PS Example 4; Fig 1; 50pp; French.
XX
CC This is the amino acid sequence of the protein encoded by the open
CC reading frame 2 (ORF-2) from the BamHI I fragment from the herpes virus
CC of turkeys (HVT). The protein has homology to the 185 amino acid protein
CC encoded by the FHV-1 gene 3 ORF. The DNA fragment was isolated from HVT
CC strain FCI26 (J. Gen. Virol., 70:1789-1804 (1989)) and is used as a basis
CC for generating live recombinant viral vaccines using the herpes virus DNA
CC sequences as vectors. The fragment contains 3 intergenic regions into
CC which genes encoding heterologous proteins to be expressed can be
CC inserted, preferably under control of the cytomegalovirus immediate early
CC (CMV-IE) promoter. The heterologous proteins can be used to vaccinate
CC poultry against Gunbore disease (caused by IBDV), Newcastle disease,
CC Marek's disease, infectious bronchitis, infectious laryngotracheitis and
CC avian anaemia.
XX
SQ Sequence 185 AA:

Query Match 96.6%; Score 28; DB 17; Length 185;
Best Local Similarity 83.3%; Pred. No. 33;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVVKGK 6
11:111
Db 113 mvikgk 118

RESULT 2
W97389
ID W97389 standard; Protein; 415 AA.
XX
AC W97389;
XX
DT 14-MAY-1999 (first entry)
XX
DE Protein encoded by the aroA ORF.
XX
XX 5-enolpyruvylshikimate-3-phosphate synthase; aroA;
XX aromatic amino acid synthetase; vaccine; bacterial infection;
XX otilis media; conjunctivitis; pneumonia; bacteremia; sinusitis;
XX pleural empyema; endocarditis; meningitis.
XX
OS Streptococcus pneumoniae.
XX
FH Key Location/Qualifiers
FT Masc-difference 1.415 /note="Xaa- not specified"
XX
XX US5883239-A.
XX
XX 16-MAR-1999.
XX
XX 18-JUL-1997; 97US-0896345.
XX
XX 18-JUL-1997; 97US-0896345.
XX
XX 18-JUL-1997; 97US-0896345.
XX
XX (SMIK) SMITHKLINE BEECHAM CORP.
XX
XX Brown JR, Chalke AF, Payne DJ, Shilling LK, Traini CM;
XX
XX WPI: 1999-214154/18.
XX
XX N-PSDB; X15931.
XX
XX New polynucleotides and polypeptides of the
XX 5-enolpyruvylshikimate-3-phosphate synthase family - useful in
XX the creation of a vaccine against, and diagnosis and treatment of
XX Streptococcal infection, especially in meningitis

XX Disclosure; Columns 11-12; 18pp; English.
XX
XX The present sequence is encoded by the 5-enolpyruvylshikimate-3-phosphate
XX synthase (aroA) open reading frame (ORF). The aroA product converts
XX shikimate-3-phosphate to 5-enolpyruvylshikimate-3-phosphate. Inhibition
XX of this reaction prevents the synthesis of aromatic amino acids,
XX p-aminobenzoate acid (a precursor of folate) and ubiquinone. The
XX polypeptides or their variants are useful for assessing aroA expression
XX and genetic variation, the creation of a vaccine and the diagnosis and
XX treatment of bacterial infection, especially S. pneumoniae and the
XX pleural empyema, endocarditis and especially meningitis.
XX
SQ Sequence 415 AA:

Query Match 93.18; Score 27; DB 20; Length 415;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVVKGK 6
11:111
Db 353 mvikgk 358

RESULT 3
Y56505
ID Y56505 standard; Protein; 415 AA.
XX
AC Y56505;
XX
DT 15-FEB-2000 (first entry)
XX
DE Streptococcus pneumoniae aroA ORF protein sequence SEQ ID NO:4.
XX
XX Streptococcus pneumoniae; 5-enolpyruvylshikimate-3-phosphate synthase;
XX aroA; antimicrobial; antibacterial; vaccine; immune response; infection;
XX diagnosis; detection.
XX
OS Streptococcus pneumoniae.
XX
XX CA2237786-A.
XX
XX 18-JAN-1999.
XX
XX 15-JUL-1998; 98CA-2237786.
XX
XX 18-JUL-1997; 97US-0896354.
XX
XX (SMIK) SMITHKLINE BEECHAM CORP.
XX (SMIK) SMITHKLINE BEECHAM PLC.
XX
XX Payne DJ, Chalke AF, Shilling LK, Traini CM, Brown JR;
XX
XX WPI: 2000-000099/01.
XX
XX N-PSDB; Z38852.
XX
XX Polynucleotide sequences encoding Streptococcal
XX 5-enolpyruvylshikimate-3-phosphate synthase enzymes useful for
XX screening candidate antibacterial compounds -
XX
XX Disclosure; Page 16; 43pp; English.
XX
XX The present sequence is 5-enolpyruvylshikimate-3-phosphate synthase
XX (aroA) isolated from Streptococcus pneumoniae. AroA polynucleotides and
XX protein may be used in the diagnosis, prevention and treatment of
XX diseases associated with the expression of aroA. For example, the
XX polynucleotide and aroA proteins may be used to treat disorders
XX associated with decreased aroA expression. AroA polynucleotides may be
XX administered to treat diseases by rectifying mutations or deletions in a
XX patient's genome that affect the activity of aroA by expressing inactive
XX proteins or to supplement the patients own production of aroA proteins.

CC Conversely, antisense nucleic acid molecules may be administered to down
CC regulate aroA expression by binding with the cells own aroA genes and
CC preventing their expression. They may also be used to study the
CC expression and function of aroA proteins and their role in cellular
CC metabolism. The aroA proteins may then be used as antigens in the
CC production of antibodies against aroA and in assays to identify
CC modulators (agonists and antagonists) of aroA expression and activity.
CC The anti-aroA antibodies and aroA antagonists may also be used to down
CC regulate aroA expression and activity. They may be used in this way to
CC treat diseases caused by Streptococcus pneumoniae. The anti-aroA
CC antibodies may also be used as diagnostic agents for detecting the
CC presence of aroA polypeptides in samples (e.g. by enzyme linked
CC immunosorbent assay (ELISA)).

XX
XX
SQ Sequence 415 AA:

Query Match 93.1%; Score 27; DB 21; Length 415;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 MYVKCK 6
I::|||

DB 353 milkgk 358

RESULT 4
W97388 W97388 standard; Protein: 427 AA.
XX
XX W97388:
XX 14-MAY-1999 (first entry)
XX
XX 5-enolpyruvylshikimate-3-phosphate synthase designated aroA.
XX
XX 5-enolpyruvylshikimate-3-phosphate synthase; aroA;
XX aromatic amino acid synthesis; vaccine; bacterial infection;
XX otitis media; conjunctivitis; pneumonia; bacteremia; sinusitis;
XX pleural empyema; endocarditis; meningitis.
XX
XX Streptococcus pneumoniae.
XX
XX US5863239-A.
XX
XX 16-MAR-1999.
XX
XX 18-JUL-1997; 97US-0896345.
XX
XX 18-JUL-1997; 97US-0896345.
XX
XX (SMRK) SMITHKLINE BEECHAM CORP.
XX
XX Brown JR, Chalker AF, Payne DJ, Shilling LK, Traini CM;
XX
XX MPI: 1999-214154/18.
XX N-PSDB: X15930.
XX
XX New polynucleotides and polypeptides of the
XX 5-enolpyruvylshikimate-3-phosphate synthase family - useful in
XX the creation of a vaccine against, and diagnosis and treatment of
XX Streptococcal infection, especially in meningitis

PS Claim 1: Columns 9-8: 18pp; English.

XX
XX The present sequence represents a 5-enolpyruvylshikimate-3-phosphate
XX synthase designated aroA. The aroA product converts shikimate-3-phosphate
XX to 5-enolpyruvylshikimate-3-phosphate. Inhibition of this reaction
XX prevents the synthesis of aromatic amino acids, p-aminobenzoate acid (a
XX precursor of folate) and ubiquinone. The polypeptides or their variants
XX are useful for assessing aroA expression and genetic variation, the
XX creation of a vaccine and the diagnosis and treatment of bacterial
XX infection, especially S. pneumoniae and the diseases otitis media,

CC conjunctivitis, pneumonia, bacteremia, sinusitis, pleural empyema,
CC endocarditis and especially meningitis.

XX
XX
SQ Sequence 427 AA:

Query Match 93.1%; Score 27; DB 20; Length 427;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 MYVKCK 6
I::|||

DB 365 milkgk 370

RESULT 5
Y56504 Y56504 standard; Protein: 427 AA.
XX
XX Y56504:
XX 15-FEB-2000 (first entry)
XX
XX Streptococcus pneumoniae aroA protein sequence SEQ ID NO:2.
XX
XX Streptococcus pneumoniae; 5-enolpyruvylshikimate-3-phosphate synthase;
XX aroA; antimicrobial; antibacterial; vaccine; immune response; infection;
XX diagnosis; detection.
XX
XX Streptococcus pneumoniae.
XX
XX CA2237786-A.
XX
XX 18-JAN-1999.
XX
XX 15-JUL-1998; 98CA-2237786.
XX
XX 18-JUL-1997; 97US-0896354.
XX
XX (SMRK) SMITHKLINE BEECHAM CORP.
XX (SMRK) SMITHKLINE BEECHAM PLC.
XX
XX Payne DJ, Chalker AF, Shilling LK, Traini CM, Brown JR;
XX
XX MPI: 2000-000099/01.
XX N-PSDB: Z38851.
XX
XX Polynucleotide sequences encoding Streptococcal
XX 5-enolpyruvylshikimate-3-phosphate synthase enzymes useful for
XX screening candidate antibacterial compounds -

PS Claim 12: Page 13; 43pp; English.

XX
XX The present sequence is 5-enolpyruvylshikimate-3-phosphate synthase
XX (aroA) isolated from Streptococcus pneumoniae. AroA polynucleotides and
XX protein may be used in the diagnosis, prevention and treatment of
XX diseases associated with the expression of aroA. For example, the
XX polynucleotide and aroA proteins may be used to treat disorders
XX associated with decreased aroA expression. AroA polynucleotides may be
XX administered to treat diseases by rectifying mutations or deletions in a
XX patient's genome that affect the activity of aroA by expressing inactive
XX proteins or to supplement the patients own production of aroA proteins.
XX Conversely, antisense nucleic acid molecules may be administered to down
XX regulate aroA expression by binding with the cells own aroA genes and
XX preventing their expression. They may also be used to study the
XX expression and function of aroA proteins and their role in cellular
XX metabolism. The aroA proteins may then be used as antigens in the
XX production of antibodies against aroA and in assays to identify
XX modulators (agonists and antagonists) of aroA expression and activity.
XX The anti-aroA antibodies and aroA antagonists may also be used to down
XX regulate aroA expression and activity. They may be used in this way to
XX treat diseases caused by Streptococcus pneumoniae. The anti-aroA
XX antibodies may also be used as diagnostic agents for detecting the

CC presence of aroa polypeptides in samples (e.g. by enzyme linked
CC immunosorbent assay (ELISA)).

XX Sequence 427 AA;

SO

Query Match 93.1%; Score 27; DB 21; Length 427;
Best Local Similarity 66.7%; Pred. No. 1.4e+02;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 MVMGK 6
|::|||

Db 365 mlkqk 370

RESULT: 6

R78382 ID R78382 standard; peptide; 64 AA.

AC R78382;

DT 16-APR-1996 (first entry)

DE Truncated TMV movement protein.

XX Truncated; tobacco mosaic virus; TMV; movement protein; coat protein;
XX frame shift; mutation; in-frame stop codon; antigen; mammal;
KM viral vaccine; contraceptive; immune response; plant virus; insect.

XX Synthesis.

PN W09521248-A1.

PD 10-AUG-1995.

XX 10-AUG-1995.

PF 03-FEB-1995; 95MO-US01467.

XX 03-FEB-1995; 95MO-US01467.

PR 03-FEB-1994; 94US-0192477.

PA (SGRI) SCRIPPS RES INST.

PI Beachy RN, Fitchen JH;

DR WPI; 1995-283768/37.

XX Overprod. of heterologous peptide in plants via tobacco mosaic
XX virus infection - in which the coat protein gene is modified by
PT insertion of heterologous sequence, partic. for producing viral
PT antigens for vaccines

XX Claim 5; Page 98; 157pp; English.

XX This sequence represents a truncated tobacco mosaic virus (TMV)
CC movement protein. This sequence is produced by a frame shift
CC insertion of a BamHI/KpnI fragment of TMV. The nucleotides AGCT are
CC introduced to the 30 end of both strands of a HindIII site at viral
CC position 5080. This leads to the production of an in-frame stop
CC codon causing production of the 64 residue protein, rather than the
CC wild type 267 amino acid protein. The resulting protein is non-
CC functional and viruses containing it cannot spread from one plant
CC cell to another. Recombinant viruses containing this modified MP
CC may be used in the method of the invention for the overproduction
CC of heterologous proteins. The method is esp. used for the production
CC of peptides which cause an antigenic response in mammals, partic. for
CC use in viral vaccines, for generating a contraceptive immune
CC response. The infectious clones may encode coat proteins for other
CC plant viruses, or proteins conferring protection against insects and
CC diseases.

XX Sequence 64 AA;

Query Match

89.7%; Score 26; DB 16; Length 64;

Best Local Similarity 83.3%; Pred. No. 30;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MVMGK 6
|::|||

Db 3 lvmqk 8

RESULT 7

ID Y82505 standard; protein; 187 AA.

AC Y82505;

DT 12-JUL-2000 (first entry)

DE Human adenovirus type 15 fibre knob protein sequence SPQ ID NO:11.

XX Adenovirus; adenoviral; Ad; fibre knob; fibre protein; gene therapy;
KM infection; virion.

XX Human adenovirus type 15.

PN W0200015823-A1.

PD 23-MAR-2000.

PF 10-SEP-1999; 99MO-US20728.

XX 11-SEP-1998; 98US-0098851.

PR 28-MAY-1999; 99US-0136529.

PA (GENV-) GENVEC INC.

PI Wickham TJ, Kovesdi I, Roelink PM, Bruder JT;

DR WPI; 2000-271459/23.

XX Recombinant fiber protein used for creating vectors for delivering a
XX desired gene to a cell with minimal ectopic infection, comprises an
PT amino terminus of an adenoviral fiber protein and a trimerization
PT domain -

XX Example 2; Fig 1; 69pp; English.

XX The present invention describes a recombinant fibre protein (I)
CC comprising an amino terminus of an adenoviral fibre protein and a
CC trimerisation domain. The trimerisation domain comprises: (a) an
CC adenoviral fibre knob domain having a mutation affecting at least one
CC amino acid residue within the region corresponding to the Ab loop,
CC B sheet, DE loop or FG loop of the wild-type adenovirus (Ad) 5 fibre
CC protein, and where the recombinant fibre protein trimerises when
CC produced in a eukaryotic cell; or (b) an adenoviral fibre knob domain
CC having a mutation affecting at least one amino acid residue
CC corresponding to residue positions 404-406, 408, 409, 412-417, 420, 439,
CC 441, 442, 449-454, 458, 460, 462, 467, 469-472, 474-477, 482, 485,
CC 487-492, 505-512, 515, 519, 521-528, 533, 535, 537-549, 551, 553,
CC 555, 559-568, 580 or 581 of the wild-type Ad5 fibre protein, and where
CC the recombinant fibre protein trimerises when produced in a eukaryotic
CC cell. (I) are used for creating vectors which are useful in a variety
CC of gene-transfer applications, in vitro and in vivo, for delivering a
CC desired gene to a cell with minimal ectopic infection. Particularly,
CC the vectors are useful for infecting a desired cell type, an approach
CC known as alternative targeting. (I) can be used in receptor-ligand
CC assays and as adhesion proteins in vitro or in vivo. The recombinant
CC fibre protein permits more efficient production and construction of
CC safer vectors for gene transfer applications. The present sequence
CC represents an adenovirus fibre knob protein sequence, which is used
CC in the exemplification of the present invention.

XX Sequence 187 AA;

Query Match 89.7%: Score 26; DB 21; Length 187;
 Best Local Similarity 83.3%: Pred. No. 95;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MYVKGK 6
 :|||||
 Db 40 LVKKGK 45

RESULT 8

R60610 R60610 standard; Protein: 264 AA.

AC R60610;

DT 21-JUN-1995 (first entry)

DE Tobamovirus movement protein.

XX Tobamovirus; tobacco mosaic virus; vector; transcription; promoter;

KM terminator; replication enzyme; replication site; ori; coat protein.

XX Tobacco mosaic virus.

OS JP06233688-A.

XX 23-AUG-1994.

XX 29-NOV-1993; 93JP-0297704.

XX 01-DEC-1992; 92JP-0322100.

XX (SANY) SANKYO CO LTD.

XX WPI: 1994-319190/40.

DR N-PSDB; Q71546.

XX Improved transcriptional vector of plant virus - esp. vectors

XX tested for stability and function

XX Disclosure: Page 27-28; 36pp; Japanese.

XX A transcriptional vector comprises in a 5'-3' direction (i) a 5'

CC non-transcribed region of tobamovirus, (ii) a replicative enzyme of

CC tobamovirus, (iii) a foreign gene connecting region, and (iv) a 3'

CC non-transcribed region of tobamovirus. The vector also has a

CC transcriptional promoter region upstream of the tobamovirus

CC region and a transcriptional terminator region downstream of the

CC tobamovirus region, neither of which is recognised by the host cell,

CC a replication initiation site which functions in the host cell, and

CC a selective marker gene. This sequence is the tobamovirus

CC movement (sic) protein and the sequence encoding it was used in a

CC construct (See Q71546) which in turn was used in the construction of

CC the transcriptional vectors pL5013 and pL5020.

XX Sequence 264 AA;

XX Query Match 89.7%: Score 26; DB 15; Length 264;

XX Best Local Similarity 83.3%: Pred. No. 14e+02;

XX Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MYVKGK 6

XX :|||||

Db 3 LVKKGK 8

XX RESULT 9

XX R67755 R67755 standard; Protein: 264 AA.

XX AC R67755;

XX 20-JUL-1995 (first entry)

XX TMV P30 elicitor.

XX Cucurbit mosaic virus; tomato mosaic virus; minus-strand RNA;

KM virus resistance; disease resistance; transgenic plant;

KM cross protection; hypersensitive response; crop improvement;

KM tomato; Lycopersicon esculentum; tobacco; Nicotiana tabacum;

XX P30; elicitor; TMV; CMV.

XX Tomato mosaic virus.

XX WO9429464-A.

XX 22-DEC-1994.

XX 03-JUN-1994; 94WO-EP01817.

XX 04-JUN-1993; 93GB-0011593.

XX (SANO) SANDOZ LTD.

XX (SANO) SANDOZ PATENT GMBH.

XX (SANO) SANDOZ-ERFINDUNGEN VERW GES MBH.

XX De HAAN PT;

XX WPI: 1995-036490/05.

DR N-PSDB; Q76108.

XX DNA constructs for imparting pathogen, esp. virus, resistance to

XX plants - encodes RNA that interacts with viral RNA polymerase to

XX generate an eliciting agent.

XX Disclosure: Page 36; 50pp; English.

XX A chimeric cucumber mosaic virus RNA-3 (given in Q76108) has the

CC coat protein gene replaced by a gene encoding an elicitor.

CC TOMV P30, having the sequence given in R67755. The

CC construct elicits a minus-sense RNA that interacts with the

CC RNA-dependent RNA-polymerase of an invading virus, thus conferring

CC virus-resistance on a host plant, e.g. tobacco, tomato.

XX Sequence 264 AA;

XX Query Match 89.7%: Score 26; DB 16; Length 264;

XX Best Local Similarity 83.3%: Pred. No. 1.4e+02;

XX Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Oy 1 MYVKGK 6

XX :|||||

Db 3 LVKKGK 8

XX RESULT 10

XX R78388 R78388 standard; Protein: 268 AA.

XX AC R78388;

XX DT 17-APR-1996 (first entry)

XX Tobacco mosaic virus movement protein.

XX Truncated; tobacco mosaic virus; TMV; movement protein; coat protein;

XX frame shift; mutation; in-frame stop codon; antigen; mammal; primer;

XX viral vaccine; contraceptive; immune response; plant virus; insect;

XX polymerase chain reaction; PCR; amplify.

XX Tobacco mosaic virus.

XX WO9521248-A1.

[illegible]

```

XX      (BIOS-) BIOSOURCE TECHNOLOGIES INC.
XX
XX      Turpen TH;
XX
XX      WPI; 1998-530986/45.
XX
XX      Plant transformation system - comprises replicase-deficient
XX      tobamovirus vector and movement-protein deficient helper virus
XX
XX      Disclosure: Columns 19-22; 20pp; English.
XX
XX      The present sequence appears in the specification, which describes a
XX      plant transformation system which comprises a replicon and a helper
XX      virus. The replicon comprises a tobamovirus replication origin, at least
XX      one non-tobamovirus gene encoding a non-tobamovirus protein and a gene
XX      encoding a tobamovirus movement protein but lacks a gene encoding a
XX      tobamovirus replicase. The helper virus is a tobamovirus that has a gene
XX      encoding a tobamovirus replicase but lacks a functional gene encoding the
XX      tobamovirus movement protein. DNA encoding the replicon is integrated as
XX      a transgene into a chromosome of a tobamovirus-susceptible plant cell.
XX      The plant transformation system is useful for producing the
XX      non-tobamovirus protein by integrating the replicon DNA into a chromosome
XX      of a plant cell and infecting the plant cell with the helper virus so
XX      that the protein is expressed and accumulates in the plant cell. The
XX      protein produced can be a regulatory enzyme or a structural, regulatory
XX      or therapeutic protein, e.g. an interleukin or colony-stimulating factor.
XX
XX      Sequence    268 AA;
XX
XX      Query Match          89.7%; Score 26; DB 19; Length 268;
XX      Best Local Similarity 83.3%; Pred. No. 1.4e+02;
XX      Matches   5; Conservative   1; Mismatches   0; Indels   0; Gaps   0;
OY      1 MYVKGK 6
        : |||||
        3 lrvkxgk 8
Db
RESULT 12
Y68661 Y68661 standard; Protein; 339 AA.
XX
XX      Y68661;
XX
XX      05-MAY-2000 (first entry)
XX
XX      Amino acid sequence of the fiber protein of Adenovirus serotype 39.
XX
XX      Chimaeric adenovirus; gene therapy; antigenicity; fiber protein;
XX      serotype 39; penton protein; hexon protein.
XX
XX      Adenovirus.
XX
XX      WO200003029-A2.
XX
XX      20-JAN-2000.
XX
XX      08-JUL-1999; 99WO-NL00436.
XX
XX      08-JUL-1998; 98EP-0202297.
XX
XX      (INTR-) INTROGENE BV.
XX
XX      Havenga M, Vogels R, Bout A;
XX
XX      WPI; 2000-171149/15.
XX
XX      New chimaeric adenoviruses containing a genome derived from different
XX      adenovirus serotypes, useful in gene therapy -
XX
XX      Example 2; Fig 7; 92pp; English.

```

XX Y68642-70 represent the amino acid sequences of the fiber proteins
CC of Adenovirus serotypes 8, 9, 13, 14, 20, 23, 24, 25, 27, 28, 29, 30,
CC 32, 33, 34, 35, 36, 37, 38, 39, 42, 43, 44, 45, 46, 47, 48, 49 and 51.
CC The proteins are used in the course of the invention to construct
CC chimeric adenoviruses with reduced antigenicity. The chimeric
CC adenoviruses comprise at least part of a fiber protein of an adenovirus
CC serotype providing the chimeric virus with a desired host range and at
CC least part of a penton or hexon protein from another, less antigenic,
CC serotype. The chimeric adenoviruses are useful for gene therapy,
CC especially where repeated delivery is required. Adenoviruses of the
CC invention are useful can be constructed to have a desired host range and
CC a diminished capability to raise neutralizing antibodies, an absence of,
CC or decreased infection of, antigen presenting cells of the immune system
CC (e.g. macrophages), and an ability to escape trapping in the liver
CC through increased target cell specificity.
CC XX

Sequence 339 AA:

Query Match 89.7%; Score 26; DB 21; Length 339;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MGVKKG 6
:|||||
Db 200 lvvkkg 205

RESULT 13
ID Y68664 Y68664 standard; Protein; 341 AA.
XX Y68664;
AC
XX
XX 05-MAY-2000 (first entry)
DT
XX
XX Amino acid sequence of the fiber protein of Adenovirus serotype 44.
DE
XX
XX Chimeric adenovirus; gene therapy; antigenicity; fiber protein;
KM serotype 44; penton protein; hexon protein.
XX
XX Adenovirus.
OS
XX
XX WO200003029-A2.
PN
XX
XX 20-JAN-2000.
PD
XX
XX 08-JUL-1999; 99WO-NL00436.
PF
XX
XX 08-JUL-1998; 98EP-0202297.
PR
XX
XX (INTR-) INTRIGENE BV.
PA
XX
XX Havenga M, Vogels R, Bout A;
PI
XX
XX WPI; 2000-171149/15.
DR
XX
XX New chimeric adenoviruses containing a genome derived from different
PT adenovirus serotypes, useful in gene therapy -
XX
XX
XX Example 2; Fig 7; 92pp; English.
PS
XX Y68642-70 represent the amino acid sequences of the fiber proteins
CC of Adenovirus serotypes 8, 9, 13, 14, 20, 23, 24, 25, 27, 28, 29, 30,
CC 32, 33, 34, 35, 36, 37, 38, 39, 42, 43, 44, 45, 46, 47, 48, 49 and 51.
CC The proteins are used in the course of the invention to construct
CC chimeric adenoviruses with reduced antigenicity. The chimeric
CC adenoviruses comprise at least part of a fiber protein of an adenovirus
CC serotype providing the chimeric virus with a desired host range and at
CC least part of a penton or hexon protein from another, less antigenic,
CC serotype. The chimeric adenoviruses are useful for gene therapy,
CC especially where repeated delivery is required. Adenoviruses of the

CC invention are useful can be constructed to have a desired host range and
CC a diminished capability to raise neutralizing antibodies, an absence of,
CC or decreased infection of, antigen presenting cells of the immune system
CC (e.g. macrophages), and an ability to escape trapping in the liver
CC through increased target cell specificity.
CC XX

Sequence 341 AA:

Query Match 89.7%; Score 26; DB 21; Length 341;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MGVKKG 6
:|||||
Db 198 lvvkkg 203

RESULT 14
ID Y68668 Y68668 standard; Protein; 343 AA.
XX Y68668;
AC
XX
XX 05-MAY-2000 (first entry)
DT
XX
XX Amino acid sequence of the fiber protein of Adenovirus serotype 48.
DE
XX
XX Chimeric adenovirus; gene therapy; antigenicity; fiber protein;
KM serotype 48; penton protein; hexon protein.
XX
XX Adenovirus.
OS
XX
XX WO200003029-A2.
PN
XX
XX 20-JAN-2000.
PD
XX
XX 08-JUL-1999; 99WO-NL00436.
PF
XX
XX 08-JUL-1998; 98EP-0202297.
PR
XX
XX (INTR-) INTRIGENE BV.
PA
XX
XX Havenga M, Vogels R, Bout A;
PI
XX
XX WPI; 2000-171149/15.
DR
XX
XX New chimeric adenoviruses containing a genome derived from different
PT adenovirus serotypes, useful in gene therapy -
XX
XX
XX Example 2; Fig 7; 92pp; English.
PS
XX Y68642-70 represent the amino acid sequences of the fiber proteins
CC of Adenovirus serotypes 8, 9, 13, 14, 20, 23, 24, 25, 27, 28, 29, 30,
CC 32, 33, 34, 35, 36, 37, 38, 39, 42, 43, 44, 45, 46, 47, 48, 49 and 51.
CC The proteins are used in the course of the invention to construct
CC chimeric adenoviruses with reduced antigenicity. The chimeric
CC adenoviruses comprise at least part of a fiber protein of an adenovirus
CC serotype providing the chimeric virus with a desired host range and at
CC least part of a penton or hexon protein from another, less antigenic,
CC serotype. The chimeric adenoviruses are useful for gene therapy,
CC especially where repeated delivery is required. Adenoviruses of the
CC invention are useful can be constructed to have a desired host range and
CC a diminished capability to raise neutralizing antibodies, an absence of,
CC or decreased infection of, antigen presenting cells of the immune system
CC (e.g. macrophages), and an ability to escape trapping in the liver
CC through increased target cell specificity.
CC XX

Sequence 343 AA:

Query Match 89.7%; Score 26; DB 21; Length 343;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MVRGK 6
:|||||
Db 199 lvrkgk 204

RESULT 15

ID Y68665
Y68665 standard; Protein; 345 AA.

AC Y68665;

DT 05-MAY-2000 (first entry)

DE Amino acid sequence of the fiber protein of Adenovirus serotype 45.

KM Chimeric adenovirus; gene therapy; antigenicity; fiber protein;
serotype 45; penton protein; hexon protein.

OS Adenovirus.

PN W6300003029-A2.

PD 20-JAN-2000.

PF 08-JUL-1999; 99WO-NL00436.

PR 08-JUL-1998; 98EP-0202297.

PA (PVR-) INTROGENE BV.

PI Hawenga M, Vogels R, Bout A;

DR WPI 2000-171149/15.

PT New chimeric adenoviruses containing a genome derived from different
adenovirus serotypes, useful in gene therapy -

PS Example 2: Fig 7; 92pp: English.

XX Y68642-70 represent the amino acid sequences of the fiber proteins
of Adenovirus serotypes 8, 9, 13, 14, 20, 23, 24, 25, 27, 28, 29, 30,
32, 33, 34, 35, 36, 37, 38, 39, 42, 43, 44, 45, 46, 47, 48, 49 and 51.
XX The proteins are used in the course of the invention to construct
XX chimeric adenoviruses with reduced antigenicity. The chimeric
XX adenoviruses comprise at least part of a fiber protein of an adenovirus
XX serotype providing the chimeric virus with a desired host range and at
XX least part of a penton or hexon protein from another, less antigenic,
XX serotype. The chimeric adenoviruses are useful for gene therapy,
XX especially where repeated delivery is required. Adenoviruses of the
XX invention are useful can be constructed to have a desired host range and
XX a diminished capability to raise neutralizing antibodies, an absence of,
XX or decreased infection of, antigen presenting cells of the immune system
XX (e.g. macrophages), and an ability to escape trapping in the liver
XX through increased target cell specificity.

SO Sequence 345 AA;

Query Match 89.7%; Score 26; DB 21; Length 345;

Best Local Similarity 83.3%; Pred. No. 1.8e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MVRGK 6
:|||||

Db 203 lvrkgk 208

Search completed: December 14, 2000, 10:11:11
Job time: 45 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 14, 2000, 10:11:40 : Search time 27.22 seconds
(without alignments)
3.695 Million cell updates/sec

Title: US-09-051-034a-13

Perfect score: 29

Sequence: 1 MVVKGK 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 164575 seqs, 16761186 residues

Total number of hits satisfying chosen parameters: 164575

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents_AA:*
1: /cgn2_6/prodata/2/1aa/5a_COMB.pep.*
2: /cgn2_6/prodata/2/1aa/5b_COMB.pep.*
3: /cgn2_6/prodata/2/1aa/6_COMB.pep.*
4: /cgn2_6/prodata/2/1aa/PCTUS_COMB.pep.*
5: /cgn2_6/prodata/2/1aa/Backfile91.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	27	93.1	415	2	US-08-896-345-4
2	27	93.1	427	2	US-08-896-345-2
3	26	89.7	64	2	US-08-887-559-3
4	26	89.7	264	2	US-08-553-6198-7
5	26	89.7	268	2	US-08-176-4148-2
6	26	89.7	268	2	US-08-687-559-25
7	26	89.7	472	1	US-08-687-559-25
8	26	89.7	472	1	US-08-749-903-1
9	26	89.7	472	1	US-08-749-903-3
10	26	89.7	472	1	US-08-749-903-4
11	26	89.7	1049	3	US-08-749-903-5
12	26	89.7	1049	3	US-08-772-270A-11
13	26	89.7	1244	4	PCT-US93-10500-2
14	25	86.2	93	3	US-08-946-329A-95
15	25	86.2	99	3	US-08-946-329A-96
16	25	86.2	128	2	US-08-946-528-5
17	25	86.2	128	2	US-08-946-528-9
18	25	86.2	336	4	PCT-US92-10421-2
19	24	82.8	435	5	5268463-9
20	24	82.8	435	5	5432081-10
21	24	82.8	714	2	US-08-990-114-3
22	24	82.8	1152	2	US-08-476-062A-43
23	24	82.8	1152	4	PCT-US96-01314-43
24	24	82.8	1152	5	5424399-2
25	24	82.8	1153	1	US-08-173-497-3
26	24	82.8	1153	1	US-08-286-889-3
27	24	82.8	1153	1	US-08-485-618-3
28	24	82.8	1153	1	US-08-362-652-3
					Sequence 3, Appl1

29	24	82.8	1153	2	US-08-482-293A-3	Sequence 3, Appl1
30	24	82.8	1153	2	US-08-943-363-3	Sequence 3, Appl1
31	24	82.8	1786	2	US-08-477-451-16	Sequence 16, Appl1
32	23	79.3	10	1	US-08-594-447-12	Sequence 12, Appl1
33	23	79.3	10	1	US-08-541-964-11	Sequence 11, Appl1
34	23	79.3	10	2	US-08-665-647-26	Sequence 26, Appl1
35	23	79.3	49	1	US-08-221-285-4	Sequence 4, Appl1
36	23	79.3	264	2	US-08-468-5768-14	Sequence 14, Appl1
37	23	79.3	264	2	US-08-468-5768-14	Sequence 14, Appl1
38	23	79.3	264	2	US-08-468-5768-14	Sequence 14, Appl1
39	23	79.3	368	1	US-08-423-3998-35	Sequence 35, Appl1
40	23	79.3	379	1	US-07-723-002C-6	Sequence 6, Appl1
41	23	79.3	423	1	US-08-445-746-2	Sequence 2, Appl1
42	23	79.3	423	3	US-08-008-722-2	Sequence 2, Appl1
43	23	79.3	547	2	US-08-467-822-35	Sequence 35, Appl1
44	23	79.3	573	2	US-08-706-209-1	Sequence 1, Appl1
45	23	79.3	573	3	US-08-981-787-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-08-896-345-4
: Sequence 4, Application US/08896345
: Patent No. 5883239
: GENERAL INFORMATION:
: APPLICANT: Chalket, Allison Frances
: APPLICANT: Traini,
: TITLE OF INVENTION: NOVEL arca
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dechert Price & Rhoads
: STREET: 997 Lenox Drive, Building 3, Suite 210
: CITY: Lawrenceville
: STATE: NJ
: COUNTRY: USA
: ZIP: 08543
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FASTSEQ for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/896,345
: FILING DATE:
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Bloom, Allen
: REGISTRATION NUMBER: 29,135
: REFERENCE/DOCKET NUMBER: P50029-01
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 609-520-3214
: TELEFAX: 609-520-3259
: TELEX:
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 415 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-896-345-4

Query Match 93.1%, Score 27, DB 2, Length 415:
Best Local Similarity 66.7%, Pred. No. 96:
Matches 4, Conservative 2, Mismatches 0, Indels 0, Gaps 0,
OY 1 MVVKGK 6
|::|||

Db 353 MIIRK 358

RESULT 4 2

US-08-896-345-2

Sequence 2, Application US/08896345

Patent No. 5883239

GENERAL INFORMATION:

APPLICANT: Chaiker, Alison Frances

APPLICANT: Trautl, David

TITLE OF INVENTION: NOVEL AREA

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

STREET: 997 Lenox Drive, Building 3, Suite 210

CITY: Lawrenceville

STATE: NJ

COUNTRY: USA

ZIP: 08543

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/896,345

FILING DATE:

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER:

ATTORNEY/AGENT INFORMATION:

NAME: Bloom, Allen

REGISTRATION NUMBER: 29,135

REFERENCE/DOCKET NUMBER: P50029-01

TELECOMMUNICATION INFORMATION:

TELEPHONE: 609-520-3214

TELEFAX: 609-520-3259

TELEX:

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 427 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

US-08-896-345-2

QY 1 MYVKG 6

Db 365 MIIRK 370

Query Match 93.1%; Score 27; DB 2; Length 427;

Best Local Similarity 66.7%; Pred. No. 99;

Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

RESULT 3

US-08-887-559-3

Sequence 3, Application US/08687559

Patent No. 5955647

GENERAL INFORMATION:

APPLICANT: Fitcher, John H.

APPLICANT: Beachy, Roger N.

TITLE OF INVENTION: METHOD FOR USING TOBACCO MOSAIC VIRUS TO

NUMBER OF SEQUENCES: 30

CORRESPONDENCE ADDRESS:

STREET: 4225 Executive Square, Suite 1400

CITY: La Jolla

STATE: CA

COUNTRY: US

ZIP: 92037

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: Windows95

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/687,559

FILING DATE: No. 5955647ember 18, 1996

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/01467

FILING DATE: 03-FEB-1995

ATTORNEY/AGENT INFORMATION:

NAME: Leary, June M.

REGISTRATION NUMBER: 31,238

REFERENCE/DOCKET NUMBER: 07302/011001

TELECOMMUNICATION INFORMATION:

TELEPHONE: 619-678-5070

TELEFAX: 619-678-5099

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 64 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

IMMEDIATE SOURCE:

CLONE: TMV MP

US-08-687-559-3

QY 1 MYVKG 6

Db 3 LVVKG 8

Query Match 89.7%; Score 26; DB 2; Length 64;

Best Local Similarity 83.3%; Pred. No. 28;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

RESULT 4

US-08-553-619B-7

Sequence 7, Application US/08553619B

Patent No. 5919705

GENERAL INFORMATION:

APPLICANT: Dehaan, Petrus T.

TITLE OF INVENTION: Virus Resistant Plants

NUMBER OF SEQUENCES: 30

CORRESPONDENCE ADDRESS:

STREET: 975 California Avenue

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94304

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/553,619B

FILING DATE: December 1, 1995

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Marcus-Wyner, Lynn

REGISTRATION NUMBER: 34,869

REFERENCE/DOCKET NUMBER: 137-1082/PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: 415/354-3588

TELEFAX: 415/857-1125

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 264 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-553-619B-7

Query Match 89.7% Score 26; DB 2; Length 264;
Best Local Similarity 83.3% Pred. No. 1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVKGK 6
:|||||
DB 3 LVVKGK 8

RESULT 5
US-08-176-414B-2
; Sequence 2, Application US/08176414B
; Patent No. 5811653
; GENERAL INFORMATION:
; APPLICANT: Turpen, Thomas H.
; TITLE OF INVENTION: VIRAL AMPLIFICATION OF RECOMBINANT
; TITLE OF INVENTION: MESSENGER RNA IN TRANSGENIC PLANTS
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Howrey & Simon
; STREET: 1299 Pennsylvania Avenue, N.W.
; CITY: Washington D.C.
; STATE: District of Columbia
; COUNTRY: United States of America
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/176,414B
; FILING DATE: 29-Dec-1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Halliun, Albert P.
; REGISTRATION NUMBER: 25,227
; REFERENCE/DOCKET NUMBER: 00801.0038.999
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-463-8109
; TELEFAX: 650-463-8400
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 268 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-176-414B-2

Query Match 89.7% Score 26; DB 2; Length 268;
Best Local Similarity 83.3% Pred. No. 1.1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVKGK 6
:|||||
DB 3 LVVKGK 8

RESULT 6
US-08-687-559-25
; Sequence 25, Application US/08687559
; Patent No. 5955647
; GENERAL INFORMATION:
; APPLICANT: Fitchen, John H.

APPLICANT: Beachy, Roger N.
; TITLE OF INVENTION: METHOD FOR USING TOBACCO MOSAIC VIRUS TO
; TITLE OF INVENTION: OVERPRODUCE PEPTIDES AND PROTEINS
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/687,559
; FILING DATE: No. 5955647ember 18, 1996
; CLASSIFICATION: 424

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/01467
; FILING DATE: 03-FEB-1995

ATTORNEY/AGENT INFORMATION:
; NAME: Learn, June M.

REGISTRATION NUMBER: 31,238
; REFERENCE/DOCKET NUMBER: 07302/011001

TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-678-5070
; TELEFAX: 619-678-5099

INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 268 amino acids
; TYPE: amino acid

TOPOLOGY: linear
; MOLECULE TYPE: protein
; IMMEDIATE SOURCE:
; CLONE: TMV MPWT

US-08-687-559-25

Query Match 89.7% Score 26; DB 2; Length 268;
Best Local Similarity 83.3% Pred. No. 1.1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVKGK 6
:|||||
DB 3 LVVKGK 8

RESULT 7
US-08-749-903-1
; Sequence 1, Application US/08749903
; Patent No. 5759812
; GENERAL INFORMATION:
; APPLICANT: Bandman, Olga

APPLICANT: Hawkins, Phillip R.
; TITLE OF INVENTION: NOVEL HUMAN SILENIUM-BINDING PROTEIN
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: INCYTE PHARMACEUTICALS, INC.

STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: US
; ZIP: 94304

COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/749,903

;; FILING DATE: Filed Herewith
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER:
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Billings, Lucy J.
;; REGISTRATION NUMBER: 36,749
;; REFERENCE/DOCKET NUMBER: PF-0163 US
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 415-855-0555
;; TELEFAX: 415-845-4166
;; INFORMATION FOR SEQ ID NO: 1:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 472 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: No. 5759812e
;; IMMEDIATE SOURCE:
;; CLONE: 989953
US-08-749-903-1

Query Match 89.7%; Score 26; DB 1; Length 472;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MVMK 6
Db 376 LVK 381

RESULT 8
US-08-749-903-3
Sequence 3, Application US/08749903
Patent No. 5759812
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Hawkins, Phillip R.
TITLE OF INVENTION: NOVEL HUMAN SELENIUM-BINDING PROTEIN
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 Porter Drive.
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/749,903
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0163 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: No. 5759812e

;; IMMEDIATE SOURCE:
;; LIBRARY: GenBank
;; CLONE: 374792
US-08-749-903-3

Query Match 89.7%; Score 26; DB 1; Length 472;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MVMK 6
Db 376 LVK 381

RESULT 9
US-08-749-903-4
Sequence 4, Application US/08749903
Patent No. 5759812
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Hawkins, Phillip R.
TITLE OF INVENTION: NOVEL HUMAN SELENIUM-BINDING PROTEIN
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/749,903
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0163 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 227630
US-08-749-903-4

Query Match 89.7%; Score 26; DB 1; Length 472;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MVMK 6
Db 376 LVK 381

RESULT 10
US-08-749-903-5

Sequence 5, Application US/08749903
Patent No. 5759812
GENERAL INFORMATION:
APPLICANT: Bandman, Olga
APPLICANT: Hawkins, Phillip R.
TITLE OF INVENTION: NOVEL HUMAN SELENIUM-BINDING PROTEIN
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: US
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/749,903
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0163 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415-855-0555
TELEFAX: 415-845-4166
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 472 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
IMMEDIATE SOURCE:
LIBRARY: GenBank
CLONE: 298710
US-08-749-903-5

Query Match 89.7%; Score 26; DB 1; Length 472;
Best Local Similarity 83.3%; Pred. No. 1.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMYKKG 6
DB 376 LYYKKG 381

RESULT 11
US-08-772-270A-11
Sequence 11, Application US/08772270A
Patent No. 6019984
GENERAL INFORMATION:
APPLICANT: Macinnes, Janet
APPLICANT: Ricciatli, Paul
APPLICANT: Mallard, Bonnie
APPLICANT: Rosengal, Soeren
TITLE OF INVENTION: NOVEL BACTERIAL PREPARATIONS, METHOD FOR
NUMBER OF SEQUENCES: 14
PRODUCING SAME, AND THEIR USE AS VACCINES
CORRESPONDENCE ADDRESS:
ADDRESSEE: Bereskin & Parr
STREET: 40 King Street West
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5H 3Y2

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/772,270A
FILING DATE: December 23, 1996
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Gravelle, Micheline
REGISTRATION NUMBER: 40,261
REFERENCE/DOCKET NUMBER: 6580-81
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 364-7311
TELEFAX: (416) 361-1398
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 1049 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Actinobacillus pleuropneumoniae
US-08-772-270A-11

Query Match 89.7%; Score 26; DB 3; Length 1049;
Best Local Similarity 83.3%; Pred. No. 3.8e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMYKKG 6
DB 578 LYYKKG 583

RESULT 12
PCT-US93-10500-2
Sequence 2, Application PC/TUS9310500
GENERAL INFORMATION:
APPLICANT: Chang, Yung-Fu
TITLE OF INVENTION: Recombinant Vaccine For Procline
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Alan S. Korman
STREET: 1600 Empire Tower
CITY: Buffalo
STATE: New York
COUNTRY: U.S.A.
ZIP: 14202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/10500
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/972,229
FILING DATE: 05-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: Korman, Alan S.
REGISTRATION NUMBER: 33,932
REFERENCE/DOCKET NUMBER: 19603/00001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 716-853-8104
TELEFAX: 716-853-8109
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:

LENGTH: 1244 amino acids
TYPE: amino acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYBOTHEICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: Actinobacillus pleuropneumoniae
STRAIN: Serotypes 2, 3, 4, 6 and 8
INDIVIDUAL ISOLATE: Swine
CELL TYPE: Gram negative bacterium
PCT-US93-10500-2

Query Match 89.7%; Score 26; DB 4; Length 1244;
Best Local Similarity 83.3%; Pred. No. 4,4e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVVKGK 6
:|||||
Db 751 LVVKGK 756

RESULT 13
US-08-946-329A-95
Sequence 95, Application US/08946329A
Patent No. 6057091
GENERAL INFORMATION:
APPLICANT: Beachy, Philip A.
APPLICANT: Porter, Jeffrey A.
TITLE OF INVENTION: NOVEL HEDGEHOG-DERIVED POLYPEPTIDES
NUMBER OF SEQUENCES: 109
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/946,329A
FILING DATE: 07-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/061,323
FILING DATE: 07-OCT-1996
APPLICATION NUMBER: 08/729,743
FILING DATE: 10-JUL-1996
APPLICATION NUMBER: 08/567,357
FILING DATE: 04-DEC-1995
APPLICATION NUMBER: 08/349,498
FILING DATE: 02-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Halle, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07265/140001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 95:
SEQUENCE CHARACTERISTICS:
LENGTH: 93 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-946-329A-95

Query Match 86.2%; Score 25; DB 3; Length 93;
Best Local Similarity 83.3%; Pred. No. 65;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVVKGK 6
:|||||
Db 8 VVKGK 13

RESULT 14
US-08-946-329A-96
Sequence 96, Application US/08946329A
Patent No. 6057091
GENERAL INFORMATION:
APPLICANT: Beachy, Philip A.
APPLICANT: Porter, Jeffrey A.
TITLE OF INVENTION: NOVEL HEDGEHOG-DERIVED POLYPEPTIDES
NUMBER OF SEQUENCES: 109
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: USA
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/946,329A
FILING DATE: 07-OCT-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/061,323
FILING DATE: 07-OCT-1996
APPLICATION NUMBER: 08/729,743
FILING DATE: 10-JUL-1996
APPLICATION NUMBER: 08/567,357
FILING DATE: 04-DEC-1995
APPLICATION NUMBER: 08/349,498
FILING DATE: 02-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Halle, Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 07265/140001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 96:
SEQUENCE CHARACTERISTICS:
LENGTH: 99 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-946-329A-96

Query Match 86.2%; Score 25; DB 3; Length 99;
Best Local Similarity 83.3%; Pred. No. 69;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVVKGK 6
:|||||
Db 8 VVKGK 13

RESULT 15
US-08-946-528-5
Sequence 5, Application US/08946528
Patent No. 5958746
GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.
APPLICANT: Bandman, Olga
APPLICANT: Lal, Preeti
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: ELECTRON TRANSPORT PROTEINS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/946,528
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0406 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 128 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: ISLITNOT01
CLONE: 2378038
US-08-946-528-5

Query Match 86.2%; Score 25; DB 2; Length 128;
Best Local Similarity 66.7%;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVKGK 6
: : : : :
Db 82 LVYKGR 87

Search completed: December 14, 2000, 10:11:41
Job time: 75 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 14, 2000, 10:12:57 ; Search time 27.57 Seconds
(without alignments)
13.811 Million cell updates/sec

Title: US-09-051-034A-13

Perfect score: 29

Sequence: 1 MGVKKG 6

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 182106 seqs, 63460219 residues

Total number of hits satisfying chosen parameters: 182106

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	29	100.0	183	1 C69474	conserved hypothet
2	29	100.0	327	2 T06602	hypothetical prote
3	28	96.6	162	2 G70413	NADH dehydrogenase
4	27	93.1	633	2 G71026	hypothetical prote
5	27	93.1	662	2 T20570	hypothetical prote
6	26	88.7	100	1 R5EC23	ribosomal protein
7	26	88.7	161	2 T37236	probable H ⁺ -transp
8	26	88.7	264	1 WMBV73	30K protein - toma
9	26	88.7	264	1 WMBVL1	30K protein - toma
10	26	88.7	264	1 WMBVL2	30K protein - toma
11	26	88.7	267	2 JCI339	movement protein -
12	26	88.7	268	1 WMM30	30K protein - toma
13	26	88.7	268	2 S26359	30K protein - toma
14	26	88.7	302	2 S40714	hypothetical prote
15	26	88.7	333	2 T05121	hypothetical prote
16	26	88.7	346	2 C81446	homolog of E. coli
17	26	88.7	367	2 S33663	fiber protein - hu
18	26	88.7	371	2 S33664	fiber protein - hu
19	26	88.7	472	2 S27878	selenium-binding p
20	26	88.7	472	2 G01872	selenium-binding p
21	26	88.7	493	2 S63060	hypothetical prote
22	26	88.7	815	2 E70021	3-hydroxycyl-CoA
23	26	88.7	1049	1 S51784	toxin III - Actino
24	26	88.7	1052	1 B49219	toxin III - Actino
25	26	88.7	1504	2 A33602	DNA-directed DNA p
26	26	88.7	1581	2 T30832	pentanucleon en
27	25	86.2	99	2 S28800	hypothetical prote
28	25	86.2	99	1 G64092	ribosomal protein
29	25	86.2	128	2 JC5821	NADH dehydrogenase

30	25	86.2	128	2 S28245	NADH dehydrogenase
31	25	86.2	138	2 F75056	hypothetical prote
32	25	86.2	151	2 G69345	nucleoside-diphosp
33	25	86.2	163	1 E69452	leucine responsive
34	25	86.2	175	2 T40068	ATP synthase chain
35	25	86.2	177	2 H82003	50S ribosomal prot
36	25	86.2	179	2 G81232	50S ribosomal prot
37	25	86.2	179	2 T34367	hypothetical prote
38	25	86.2	204	2 D69349	conserved hypothet
39	25	86.2	231	2 H64466	hypothetical prote
40	25	86.2	263	1 B64321	cell division inh
41	25	86.2	274	2 H81284	probable enoyl-lac
42	25	86.2	285	2 S29368	translational initia
43	25	86.2	329	2 A61155	isopenicillin N sy
44	25	86.2	352	2 F75099	thiamosyl transfer
45	25	86.2	372	2 A75206	hypothetical prote

ALIGNMENTS

RESULT 1
C69474
Conserved hypothetical protein AF1796 - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: C69474
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dod
., Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirschner, E
Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.
Nature 390, 364-370, 1997
A:Authors: Uitterback, T.; Cotton, M.D.; Spriggs, T.; Artlach, P.; Kaine, B.P.; Sykes,
Smith, H.O.; Moese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch
A:Reference number: A69250; MUID:96049343
A:Accession: C69474
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-183 <KLE>
A:Cross-references: GB:AE000979; GB:AE000782; NID:g2689302; PID:g2648766; TIGR:AF1796
C:Superfamily: conserved hypothetical protein MJ1247

Query Match 100.0%; Score 29; DB 1; Length 183;
Best Local Similarity 100.0%; Pred. No. 14;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGVKKG 6
DB 125 MGVKKG 130

RESULT 2
T06602
hypothetical protein F16J13.20 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 22-Oct-1999
C:Accession: T06602
R:Bevan, M.; Hilbert, H.; Braun, M.; Holzer, E.; Brandt, A.; Duesterhoeft, A.; Bancro
submitted to the Protein Sequence Database, April 1999
A:Reference number: 215789
A:Accession: T06602
A:Molecule type: DNA
A:Residues: 1-327 <BEV>
A:Cross-references: EMBL:AL049638; GSPDB:GN00062; ATSP:F16J13.20
A:Experimental source: cultivar Columbia; BAC clone F16J13
C:Genetics:
A:Gene: ATSP:F16J13.20
A:Map position: 4

Query Match 100.0%; Score 29; DB 2; Length 327;
Best Local Similarity 100.0%; Pred. No. 24;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 1 MMYGK 6
 |||||
 Db 275 MMYGK 280

RESULT 3

G70413
 NADH dehydrogenase (ubiquinone) (EC 1.6.5.3) I chain nuoJ - Aquifex aeolicus
 C:Species: Aquifex aeolicus
 C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 05-Nov-1999
 C:Accession: G70413
 R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; V.

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
 A:Reference number: A70300; MUID:9815666
 A:Accession: G70413

A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-162 <NOF>

A:Cross-references: GB:AE000734; NID:g2983733; PIDN:ANC07301.1; PID:g2983740; GB:AE00065
 A:Experimental source: strain VF5
 C:Genetics:

A:Gene: nuoJ
 C:Superfamily: NADH dehydrogenase (ubiquinone) chain 6
 C:Keywords: membrane-associated complex; NAD; oxidoreductase

Query Match 96.6%; Score 28; DB 2; Length 162;
 Best Local Similarity 83.3%; Pred. No. 21;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MMYGK 6
 |||||
 Db 108 MMYGK 113

RESULT 4

G71026
 hypothetical protein PH1507 - Pyrococcus horikoshii

C:Species: Pyrococcus horikoshii
 C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 17-Mar-2000
 C:Accession: G71026

R:Kawaguchi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekin
 M.; Onuma, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi
 DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic a
 A:Reference number: A71000; MUID:98344137
 A:Accession: G71026

A:Status: Preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-633 <KAM>

A:Cross-references: GB:AP000006; NID:g3336133; PIDN:BA030615.1; PID:d1031558; PID:g32579
 A:Experimental source: strain OT3
 A:Note: This accession replaces an interim accession for a sequence replaced by GenBank

C:Genetics:
 A:Gene: PH1507
 C:Superfamily: unassigned ATP-binding cassette proteins; ATP-binding cassette homology
 C:Keywords: ATP
 F:438-628/Domain: ATP-binding cassette homology <ABC>

Query Match 93.1%; Score 27; DB 2; Length 633;
 Best Local Similarity 66.7%; Pred. No. 1.3e+02;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 MMYGK 6
 |||||
 Db 1 MMYGK 6

RESULT 5
 T20570
 hypothetical protein F08A8.4 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
 C:Accession: T20570

R:Harris, B.
 submitted to the EMBL Data Library, October 1997

A:Reference number: Z19294
 A:Accession: T20570
 A:Status: Preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA
 A:Residues: 1-662 <WLT>
 A:Cross-references: EMBL:299710; PIDN:CA16866.1; GSPDB:GN00019; CESP:F08A8.4
 A:Experimental source: clone F08A8

C:Genetics:
 A:Gene: CESP:F08A8.4
 A:Map position: 1
 A:Insertions: 38/3; 102/2; 156/1; 327/2; 415/3; 526/3; 618/3
 C:Superfamily: acyl-CoA oxidase

Query Match 93.1%; Score 27; DB 2; Length 662;
 Best Local Similarity 66.7%; Pred. No. 1.4e+02;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 MMYGK 6
 |||||
 Db 205 MMYGK 210

RESULT 6

R5EC23
 ribosomal protein L23 [validated] - Escherichia coli

C:Species: Escherichia coli
 C:Date: 28-Feb-1980 #sequence_revision 31-Oct-1997 #text_change 28-Jan-2000
 C:Accession: A65125; D23129; A02814
 R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.

.A.; Rose, D.J.; Mau, B.; Shao, Y.
 Science 277, 1453-1462, 1997
 A:Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A64720; MUID:97426617
 A:Accession: A65125
 A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA
 A:Residues: 1-100 <BLAT>
 A:Cross-references: GB:AE000408; GB:U00096; NID:81789694; PIDN:AC76343.1; PID:g17897

A:Experimental source: strain K-12, substrain MG1655
 R:Zurawski, G.; Zurawski, S.M.
 Nucleic Acids Res. 13, 4521-4526, 1985

A:Title: Structure of the Escherichia coli S10 ribosomal protein operon.
 A:Reference number: A23129; MUID:85242118
 A:Accession: D23129

A:Molecule type: DNA
 A:Residues: 1-100 <ZUR>
 A:Cross-references: GB:X02613; NID:942825; PIDN:CA26462.1; PID:g534975

A:Note: translation of initiator Met is not shown; the initiator overlaps the last co
 R:Wittmann-Liebold, B.; Greuer, B.
 FEBS Lett. 108, 69-74, 1979

A:Title: Primary structure of protein L23 from the Escherichia coli ribosome.
 A:Reference number: A02814; MUID:80092111
 A:Accession: A02814

A:Molecule type: protein
 A:Residues: 1-79, 81-100 <WIT>
 A:Experimental source: strain K12

R:Arnold, R.J.; Reilly, J.P.
 Anal. Biochem. 269, 105-112, 1999
 A:Title: Observation of Escherichia coli ribosomal proteins and their posttranslation

A:Reference number: A59071; MUID:99196679
 A:Contents: annotation; mass spectrographic analysis
 A:Note: no post-translational modifications were observed in mass spectrographic anal

C:Genetics:
 A:Gene: rplW

A:Map position: 73 min
C:Complex: the ribosome is composed of the large (50S) and small (30S) subunit; the large S RNA and 22 distinct proteins
C:Complex: large subunit ribosomal proteins: L1 (PIR:R5EC1), L2 (PIR:R5EC3), L3 (PIR:R5EC11), L4 (PIR:R5EC14), L5 (PIR:R5EC15), L6 (PIR:R5EC16), L7 (PIR:R5EC17), L8 (PIR:R5EC18), L9 (PIR:R5EC19), L10 (PIR:R5EC20), L11 (PIR:R5EC21), L12 (PIR:R5EC22), L13 (PIR:R5EC23), L14 (PIR:R5EC24), L15 (PIR:R5EC25), L16 (PIR:R5EC26), L17 (PIR:R5EC27), L18 (PIR:R5EC28), L19 (PIR:R5EC29), L20 (PIR:R5EC30), L21 (PIR:R5EC31), L22 (PIR:R5EC32), L23 (PIR:R5EC33), L24 (PIR:R5EC34), L25 (PIR:R5EC35), L26 (PIR:R5EC36), L27 (PIR:R5EC37), L28 (PIR:R5EC38), L29 (PIR:R5EC39), L30 (PIR:R5EC40), L31 (PIR:R5EC41), L32 (PIR:R5EC42), L33 (PIR:R5EC43), L34 (PIR:R5EC44), L35 (PIR:R5EC45), L36 (PIR:R5EC46), L37 (PIR:R5EC47), L38 (PIR:R5EC48), L39 (PIR:R5EC49), L40 (PIR:R5EC50), L41 (PIR:R5EC51), L42 (PIR:R5EC52), L43 (PIR:R5EC53), L44 (PIR:R5EC54), L45 (PIR:R5EC55), L46 (PIR:R5EC56), L47 (PIR:R5EC57), L48 (PIR:R5EC58), L49 (PIR:R5EC59), L50 (PIR:R5EC60), L51 (PIR:R5EC61), L52 (PIR:R5EC62), L53 (PIR:R5EC63), L54 (PIR:R5EC64), L55 (PIR:R5EC65), L56 (PIR:R5EC66), L57 (PIR:R5EC67), L58 (PIR:R5EC68), L59 (PIR:R5EC69), L60 (PIR:R5EC70), L61 (PIR:R5EC71), L62 (PIR:R5EC72), L63 (PIR:R5EC73), L64 (PIR:R5EC74), L65 (PIR:R5EC75), L66 (PIR:R5EC76), L67 (PIR:R5EC77), L68 (PIR:R5EC78), L69 (PIR:R5EC79), L70 (PIR:R5EC80), L71 (PIR:R5EC81), L72 (PIR:R5EC82), L73 (PIR:R5EC83), L74 (PIR:R5EC84), L75 (PIR:R5EC85), L76 (PIR:R5EC86), L77 (PIR:R5EC87), L78 (PIR:R5EC88), L79 (PIR:R5EC89), L80 (PIR:R5EC90), L81 (PIR:R5EC91), L82 (PIR:R5EC92), L83 (PIR:R5EC93), L84 (PIR:R5EC94), L85 (PIR:R5EC95), L86 (PIR:R5EC96), L87 (PIR:R5EC97), L88 (PIR:R5EC98), L89 (PIR:R5EC99), L90 (PIR:R5EC100), L91 (PIR:R5EC101), L92 (PIR:R5EC102), L93 (PIR:R5EC103), L94 (PIR:R5EC104), L95 (PIR:R5EC105), L96 (PIR:R5EC106), L97 (PIR:R5EC107), L98 (PIR:R5EC108), L99 (PIR:R5EC109), L100 (PIR:R5EC110)
A:Pathway: protein biosynthesis
C:Superfamily: Escherichia coli ribosomal protein L23
C:Keywords: protein biosynthesis; ribosome
F:1-100/Product: ribosomal protein L23 #status experimental <MAT>

Query Match 89.7% Score 26; DB 1; Length 100;
Best Local Similarity 83.3% Pred. No. 41;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MYVKGK 6
DB 61 LVVKGK 66

RESULT 7
T37236
Probable H⁺-transporting ATPase (EC 3.6.1.35) chain Vha3, vacuolar - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 11-May-2000
C:Accession: T37236; T37270
R:Okada, T.; Yamamoto, R.; Futai, M.
J. Biol. Chem. 272, 24387-24392, 1997
A:Title: The vha genes encode proteolipids of Caenorhabditis elegans vacuolar-type ATP
A:Reference number: 221636; MUID:97450964
A:Accession: T37236
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-161 <OKA>
A:Cross-References: EMBL:AB000918; PIDN:BA22596.1
R:Okada, T.; Yamamoto, R.; Futai, M.
J. Biol. Chem. 273, 22570-22576, 1998
A:Title: Multiple genes for vacuolar-type ATPase proteolipids in Caenorhabditis elegans:
A:Reference number: 221659; MUID:98380481
A:Accession: T37270
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-161 <OKA>
A:Cross-References: EMBL:AB009566; PIDN:BA75066.1
A:Gene: VHA-2
C:Genetics:
A:Gene: VHA-2
A:Superfamily: vacuolar H⁺-transporting ATPase 16k chain
C:Keywords: ATP; hydrogen ion transport; hydrolase; transmembrane protein

Query Match 89.7% Score 26; DB 2; Length 161;
Best Local Similarity 83.3% Pred. No. 64;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MYVKGK 6
DB 77 MYLKGK 82

RESULT 8
WMBVT3
30k protein - tomato mosaic virus (strain L)
C:Species: tomato mosaic virus
A:Note: host (tomato)
C>Date: 18-Apr-1984 #sequence_revision 18-Apr-1984 #text_change 16-Feb-1997
C:Accession: A04182
R:Takamatsu, N.; Ohno, T.; Meshi, T.; Okada, Y.
Nucleic Acids Res. 11, 3767-3778, 1983
A:Title: Molecular cloning and nucleotide sequence of the 30k and the coat protein cistr
A:Reference number: A93473; MUID:83220776
A:Accession: A04182

A:Molecule type: genomic RNA
A:Residues: 1-264 <TAK>
C:Superfamily: tobnavirus 30k protein
C:Keywords: DNA binding

Query Match 89.7% Score 26; DB 1; Length 264;
Best Local Similarity 83.3% Pred. No. 1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MYVKGK 6
DB 3 LVVKGK 8

RESULT 9
WMBVL1
30k protein - tomato mosaic virus (strain L11A)
N:Alternate names: transport protein
C:Species: tomato mosaic virus
C>Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Feb-1997
C:Accession: J01456
R:Calder, V.L.; Palukaitis, P.
J. Gen. Virol. 73, 165-168, 1992
A:Title: Nucleotide sequence analysis of the movement genes of resistance breaking st
A:Reference number: J01456; MUID:92113565
A:Accession: J01456
A:Molecule type: genomic RNA
A:Residues: 1-264 <CAL>
A:Note: the authors translated the codons TGC for residue 68 as Cys, AGT for residue
C:Comment: This protein is involved in cell-to-cell transport of the virus.
C:Superfamily: tobnavirus 30k protein
C:Keywords: DNA binding; transport protein

Query Match 89.7% Score 26; DB 1; Length 264;
Best Local Similarity 83.3% Pred. No. 1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MYVKGK 6
DB 3 LVVKGK 8

RESULT 10
WMBVL2
30k protein - tomato mosaic virus (strain L11)
N:Alternate names: transport protein
C:Species: tomato mosaic virus
C>Date: 31-Dec-1992 #sequence_revision 31-Dec-1992 #text_change 16-Feb-1997
C:Accession: J01457
R:Calder, V.L.; Palukaitis, P.
J. Gen. Virol. 73, 165-168, 1992
A:Title: Nucleotide sequence analysis of the movement genes of resistance breaking st
A:Reference number: J01456; MUID:92113565
A:Accession: J01457
A:Molecule type: genomic RNA
A:Residues: 1-264 <CAL>
A:Note: the authors translated the codon TGC for residue 68 as Cys
C:Comment: This protein is involved in cell-to-cell transport of the virus.
C:Superfamily: tobnavirus 30k protein
C:Keywords: DNA binding; transport protein

Query Match 89.7% Score 26; DB 1; Length 264;
Best Local Similarity 83.3% Pred. No. 1e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MYVKGK 6
DB 3 LVVKGK 8

RESULT 11

JC1339

movement protein - tobacco mosaic virus

C:Species: tobacco mosaic virus, TMV

C:Date: 30-Sep-1993 #sequence_revision 21-Jan-1997 #text_change 09-May-1997

C:Accession: JC1339

R:Yu, X.; Zhu, Y.; Ying, H.; Chen, Z.

Chinese, J. Biotechnol. 12, 211-214, 1996

A:Title: Molecular cloning, nucleotide sequencing and plant transformation of a cDNA cod

A:Reference number: JC1339

A:Accession: JC1339

A:Molecule type: mRNA

A:Residues: 1-267 <YUA>

A:Experimental source: common strain

A:Note: the authors translated the codon CTT for residue 59 as Val, CAA for residue 77 a

C:Superfamily: tobnavirus 30K protein

C:Keywords: transformation

Query Match 89.7%; Score 26; DB 2; Length 267;

Best Local Similarity 83.3%; Pred. No. 1e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVMK 6

Db 3 LVK 8

RESULT 12

MTM30-

30K protein - tobacco mosaic virus

C:Species: tobacco mosaic virus, TMV

C:Date: 18-Aug-1982 #sequence_revision 18-Aug-1982 #text_change 23-Jul-1999

C:Accession: A91965; A93926; A04181

R:Meshi, T.; Ohno, T.; Okada, Y.

J. Biochem. 91, 1441-1444, 1982

A:Title: Nucleotide sequence and its character of cistron coding for the 30 K protein of

A:Reference number: A91965; MUID:82239259

A:Accession: A91965

A:Molecule type: genomic RNA

A:Residues: 1-268 <MES>

A:Cross-references: GB:V01407; NID:g62121; PIDN:CAA24686.1; PID:g62122

A:Experimental source: strain OM

R:Gelehr, P.; Lomonosoff, G.P.; Butler, P.J.G.; Akam, M.E.; Galt, M.J.; Karn, J.

Proc. Natl. Acad. Sci. U.S.A. 79, 5818-5822, 1982

A:Title: Nucleotide sequence of tobacco mosaic virus RNA.

A:Reference number: A93926; MUID:83299880

A:Accession: A93926

A:Molecule type: genomic RNA

A:Residues: 1-179, 'N', 181-224, 'N', 226-227, 'N', 229-231, 'V', 233-259, 'A', 261-268 <GOE>

A:Cross-references: GB:V01409; NID:g62125; PIDN:CAA24689.1; PID:g663265

A:Experimental source: strain Vulgare

A:Note: the sequence contains proteins 27.9K (residues 20-268), 25.3K (43-268), and 19.5

C:Superfamily: tobnavirus 30K protein

C:Keywords: DNA binding

Query Match 89.7%; Score 26; DB 1; Length 268;

Best Local Similarity 83.3%; Pred. No. 1e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVMK 6

Db 3 LVK 8

RESULT 13

S26359

30K protein - tobacco mosaic virus

C:Species: tobacco mosaic virus, TMV

C:Date: 27-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 07-May-1999

C:Accession: S26359

R:Koh, H.K.; Song, E.K.; Lee, S.Y.; Park, Y.I.; Park, W.M.

Nucleic Acids Res. 20, 5474, 1992

A:Title: Nucleotide sequence of cDNA of the tobacco mosaic virus RNA isolated in Kore

A:Reference number: S26358; MUID:93065219

A:Accession: S26359

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: mRNA

A:Residues: 1-268 <KOH>

A:Cross-references: EMBL:X68110

A:Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1992

C:Superfamily: tobnavirus 30K protein

C:Keywords: DNA binding

Query Match 89.7%; Score 26; DB 2; Length 268;

Best Local Similarity 83.3%; Pred. No. 1e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVMK 6

Db 3 LVK 8

RESULT 14

S40714

hypothetical protein R10E11.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 03-Mar-1995

C:Accession: S40714

R:Ainscough, R.; Mortimore, B.

submitted to the EMBL Data Library, December 1993

A:Reference number: S40713

A:Accession: S40714

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-302 <AIN>

A:Cross-references: EMBL:Z29095

C:Genetics:

A:Introns: 141/2; 207/1; 286/3

Query Match 89.7%; Score 26; DB 2; Length 302;

Best Local Similarity 83.3%; Pred. No. 1.2e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVMK 6

Db 218 MVLK 223

RESULT 15

T05121

hypothetical protein F7H19.80 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 23-Jul-1999

R:Bevan, M.; Peters, S.A.; van Staveren, W.; Dirkse, W.; Stiekema, W.; Bancroft, I.;

submitted to the Protein Sequence Database, July 1998

A:Reference number: Z15399

A:Accession: T05121

A:Molecule type: DNA

A:Residues: 1-333 <BEV>

A:Cross-references: EMBL:AL031018

A:Experimental source: cultivar Columbia; BAC clone F7H19

C:Genetics:

A:Map position: 4

A:Note: F7H19.80

Query Match 89.7%; Score 26; DB 2; Length 333;

Best Local Similarity 83.3%; Pred. No. 1.3e+02;

Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVMK 6

Fri Dec 15 15:27:45 2000

us-09-051-034a-13.rpr

Page 5

Db 272 MLVGR 277

Search completed: December 14, 2000, 10:13:00
Job time: 153 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 14, 2000, 10:14:58 ; Search time 15.82 Seconds
(without alignments)
12.117 Million cell updates/sec

Title: US-09-051-034A-13

Perfect score: 29
Sequence: 1 MVMK GK 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 87993 seqs, 31947931 residues

Total number of hits satisfying chosen parameters: 87993

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : SWISSPROT_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	100.0	183	1 YH96_ARCFU	O28478 archaeglob
2	29	100.0	381	1 TAL_ANASP	P48993 anabaena sp
3	27	93.1	427	1 AROA_STRPN	O95400 streptococ
4	26	89.7	100	1 RL23_ECOLI	P02424 escherichia
5	26	89.7	161	1 VAL2_CAEEL	P34546 caenorhabd
6	26	89.7	264	1 MOV_P_TMVA	P03384 tobaccos
7	26	89.7	264	1 MOV_P_TMVA	P29799 tomato mos
8	26	89.7	264	1 MOV_P_TMVA	P29800 tomato mos
9	26	89.7	268	1 MOV_P_TMVA	P03583 tobaccos
10	26	89.7	268	1 MOV_P_TMVA	P30737 tobaccos
11	26	89.7	268	1 MOV_P_TMVA	P03582 tobaccos
12	26	89.7	367	1 FIBP_ADELS	P36847 human adeno
13	26	89.7	472	1 SBI_HUMAN	Q13268 homo sapien
14	26	89.7	472	1 SBI_MOUSE	P17323 mus musculu
15	26	89.7	472	1 SBI_MOUSE	O63836 mus musculu
16	26	89.7	493	1 YNR9_YEAST	P53923 saccharomyc
17	26	89.7	1049	1 RT31_ACTPL	P55130 actinobacill
18	26	89.7	1052	1 RT32_ACTPL	P55131 actinobacill
19	26	89.7	1504	1 DP02_YEAST	P14284 saccharomyc
20	26	89.7	1581	1 ARI_PNECA	Q12659 p pentafun
21	25	86.2	87	1 Y174_BURCE	P24580 burkholderi
22	25	86.2	99	1 RL23_HAETN	P44561 haemophilus
23	25	86.2	100	1 RL23_HAETN	P55839 actinobacill
24	25	86.2	127	1 NB4M_BOVIN	O02366 bos taurus
25	25	86.2	127	1 NB4M_BOVIN	P56556 homo sapien
26	25	86.2	151	1 NDK_ARCFU	O29491 archaeglob
27	25	86.2	263	1 Y169_METJA	P09763 methanococ
28	25	86.2	285	1 IF2B_YEAST	P09664 saccharomyc
29	25	86.2	329	1 IFNS_STRGR	O54243 streptomyc
30	25	86.2	337	1 ASPQ_PSEFL	P10182 pseudomonas
31	25	86.2	362	1 ASPQ_PSEFL	O68897 pseudomonas
32	25	86.2	364	1 DP3B_MYCGE	P47247 mycoplasma
33	25	86.2	375	1 H542_YEAST	Q12329 saccharomyc

34	25	86.2	411	1 RRI_SPIOL	P29344 spinacia ol
35	25	86.2	457	1 PCAK_ACICA	O43975 actinobact
36	25	86.2	484	1 ATPB_SYNP6	P07890 synechococ
37	25	86.2	631	1 YC10_ECOLI	P45848 escherichia
38	25	86.2	787	1 RELA_MYCIE	O45640 mycobacteri
39	25	86.2	790	1 RELA_MYCIE	O50638 mycobacteri
40	25	86.2	894	1 WPA_BACSU	P54423 bacillus su
41	25	86.2	950	1 IF2_LACLA	O93764 lactococcus
42	25	86.2	1112	1 PHYE_ARATH	P43498 arabidopsis
43	25	86.2	1634	1 DPOL_METJA	O58295 methanococ
44	24	82.8	139	1 LAMP_PETMA	P33575 petromyzon
45	24	82.8	147	1 RL2B_CAEEL	P48162 caenorhabd1

ALIGNMENTS

```

RESULT 1
ID YH96_ARCFU STANDARD: PRT: 183 AA.
AC O28478:
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE HYPOTHETICAL PROTEIN AFI796.
GN AFI796.
OS Archaeoglobus fulgidus.
OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobaceae;
OC Archaeoglobus.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-VC-16 / DSM 4304 / ATCC 49558;
RX MEDLINE: 98049343.
RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,
RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.R., Peterson J.D.,
RA Richardson D.L., Kierlavage A.R., Graham D.E., Kyriades N.C.,
RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.C., Gill S.,
RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,
RA Peterson S., Reich C.I., McNeill L.K., Badger J.H., Glodex A., Zhou L.,
RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Ullrich T.,
RA Cotton M.D., Spriggs T., Arlisch P., Kalne B.P., Sykes S.M.,
RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,
RA Mason J.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,
RA Venter J.C.;
RT "The complete genome sequence of the hyperthermophilic, sulphate-
RT reducing archaeon Archaeoglobus fulgidus."
RL Nature 390:364-370(1997).
CC -!- SIMILARITY: BELONGS TO THE SIS FAMILY. YCKF SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib.ch)
CC -----
CC EMBL: AE000979; AAB89472.1; -.
CC TIGR: AFI796; -.
CC INTERPRO: IPR001347; -.
CC PFM: PFM01380; SIS: 1.
CC KW Hypothetical protein.
CC SQ SEQUENCE 183 AA: 20147 MW: FC2B8DA78E2EF14F CRC64:

```

Query Match 100.00; Score 29; DB 1; Length 183;

Best local similarity 100.00; Pred. No. 5;

Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

RESULT 2
TAL ANASP STANDARD; PRT; 381 AA.
ID P48993:
AC 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE TRANSALDOLASE (EC 2.2.1.2).
TAL.
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE; 96079511.
RA Newman J., Karakaya H., Scanlan D.J., Mann N.H.;
RT "A comparison of gene organization in the zwf region of the genomes
of the cyanobacteria Synechococcus sp. PCC 7942 and Anabaena sp. PCC
7120."
RT FEMS Microbiol. Lett. 133:187-193(1995).
CC -1- FUNCTION: TRANSALDOLASE IS IMPORTANT FOR THE BALANCE OF
CC METABOLITES IN THE PENTOSE-PHOSPHATE PATHWAY.
CC -1- CATALYTIC ACTIVITY: SEDOHEPTULOSE 7-PHOSPHATE + D-GLYCERALDEHYDE
CC 3-PHOSPHATE -> D-ERYTHROSE 4-PHOSPHATE + D-FRUCTOSE 6-PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE TRANSALDOLASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL: U33282; AA09852.1; -
DR INTERPRO; IPR001985; -
DR PFAM; PF00923; Transaldolase; 1.
DR PROSITE; PS00958; TRANSALDOLASE_2; 1.
DR PROSITE; PS01054; TRANSALDOLASE_1; 1.
KW transferase; Pentose shunt.
FT ACET_SITE 141 141 BY SIMILARITY.
SQ SEQUENCE 381 AA; 42117 MW; 32DD735130EBB84 CRC64;

Query Match 100.0%; Score 29; DB 1; Length 381;
Best Local Similarity 100.0%; Pred. No. 10;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVYK GK 6
Db 234 MVYK GK 239

RESULT 3
AROA STPN STANDARD; PRT; 427 AA.
ID AROA STPN
AC 095400;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 3-PHOSPHOSHAKIMATE 1-CARBOXYVINYLTRANSFERASE (EC 2.5.1.19) (5-
DE ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE SYNTHASE) (EPSP SYNTHASE) (EPSPS).
GN AROA.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
RN (1)
RP SEQUENCE FROM N.A.
RA Du W., Wallis N.G., Mazulla M.J., Chaliker A.F., Zhang L., Liu W.-S.,
RA Kallender H., Payne D.J.;
RT "Characterization of Streptococcus pneumoniae 5-enolpyruvylshikimate-
RT 3-phosphate synthase and its activation by monovalent cations."

```

```

RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
CC -1- CATALYTIC ACTIVITY: PHOSPHOENOLPYRUVATE + 3-PHOSPHOSHAKIMATE -
CC ORTHOPHOSPHATE + O(5)-(1-CARBOXYVINYL)-3-PHOSPHOSHAKIMATE.
CC -1- PATHWAY: SIXTH STEP IN THE BIOSYNTHESIS OF CHORISMATE WITHIN
CC THE BIOSYNTHESIS OF AROMATIC AMINO ACIDS (THE SHIKIMATE PATHWAY).
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC (PROBABLE).
CC -1- SIMILARITY: BELONGS TO THE EPSP SYNTHASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
EMBL: AF169483; AAD45819.1; -
DR INTERPRO; IPR001986; -
DR PFAM; PF00275; EPSP-synthase; 1.
DR PROSITE; PS00104; EPSP-SYNTHASE_1; 1.
DR PROSITE; PS00885; EPSP-SYNTHASE_2; 1.
KW Aromatic amino acid biosynthesis; Transferase.
SQ SEQUENCE 427 AA; 45824 MW; CE9EBD29BC641FD CRC64;

Query Match 93.1%; Score 27; DB 1; Length 427;
Best Local Similarity 66.7%; Pred. No. 34;
Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVYK GK 6
Db 365 MVYK GK 370

RESULT 4
RL23_ECOLI STANDARD; PRT; 100 AA.
ID RL23_ECOLI
AC P02424;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-MAR-1992 (Rel. 21, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE 50S RIBOSOMAL PROTEIN L23.
GN RPLW.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE; 85242118.
RA Zurawski G., Zurawski S.M.;
RT "Structure of the Escherichia coli S10 ribosomal protein operon."
RT Nucleic Acids Res. 13:4521-4526(1985).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE; 97426617.
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12."
RT Science 277:1453-1474(1997).
RN (3)
RP SEQUENCE.
RC STRAIN-K12;
RX MEDLINE; 80092111.
RA Wilmann-Liebold B., Greuer B.;
RT "Primary structure of protein L23 from the Escherichia coli
RT ribosome."
RL FEBS Lett. 108:69-74(1979).
RN (4)
RP MASS SPECTROMETRY.

```

```

RX MEDLINE; 99196679.
RA Arnold R.J., Reilly J.P.;
RT "Observation of Escherichia coli ribosomal proteins and their
RT posttranslational modifications by mass spectrometry.";
RL Anal. Biochem. 269:105-112(1999).
CC -1- FUNCTION: BINDS TO A SPECIFIC REGION ON THE 23S RRNA.
CC -1- MASS SPECTROMETRY: MW=11198.0; METHOD=MALDI.
CC -1- SIMILARITY: BELONGS TO THE 123P FAMILY OF RIBOSOMAL PROTEINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL; X02613; CA26462.1; -.
CC EMBL; U18997; AAA58115.1; -.
CC EMBL; AE000408; AAC76343.1; -.
CC PIR; D23129; R5EC23.
CC EC02DBASE; 1013.0; 6TH EDITION.
CC ECGENE; EG10883; RPLW.
CC INTERPRO; IPR001014; -.
CC PFM; PF00276; Ribosomal_L23; 1.
CC PROSITE; PS00050; RIBOSOMAL_L23; 1.
CC RIBOSOMAL protein: rRNA-binding
CC CONFLICT 80 80 MISSING (IN REF. 3).
CC FT
CC SEQUENCE 100 AA; 11199 MW; 30CD1D77C7CF9EB CAC64;
SQ
Query Match 89.7%; Score 26; DB 1; Length 100;
Best Local Similarity 83.3%; Pred. No. 15;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MVVKGK 6
Db 61 LVVKGK 66
RESULT 5
ID VAL2_CAEEL STANDARD; PRT; 161 AA.
AC P34546;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE VACUOLAR ATP SYNTHASE 16 KDA PROTEOLIPID SUBUNIT 2 (EC 3.6.1.34).
GN VHA-2 OR R10E11.2.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OC [1]
RN SEQUENCE FROM N.A.
RP MEDLINE; 97450964.
RX Oka T., Yamamoto R., Futai M.;
RT "Three vha genes encode proteolipids of Caenorhabditis elegans
RT vacuolar-type ATPase. Gene structures and preferential expression in
RT an H-shaped excretory cell and rectal cells.";
RL J. Biol. Chem. 273:24387-24392(1997).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.,
RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA Johnston L., Jones M., Kerhaw J., Kirsten J., Laister N.,
RA Larellle P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Showkhen R.,
RA Sims M., Smailson N., Smith A., Smith M., Sonhammer E., Staden R.,
RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,

```

```

RA Waterson R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA Wohldman P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RL Nature 368:32-38(1994).
RN [3]
RP REVISIONS.
RA Jones S.J.M.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: PROTON-CONDUCTING PORE FORMING SUBUNIT OF THE MEMBRANE
CC INTEGRAL V0 COMPLEX OF VACUOLAR ATPASE. V-ATPASE IS RESPONSIBLE
CC FOR ACIDIFYING A VARIETY OF INTRACELLULAR COMPARTMENTS IN
CC EUKARYOTIC CELLS.
CC -1- SUBUNIT: V-ATPASE IS AN HETEROMULTIMERIC ENZYME COMPOSED OF A
CC PERIPHERAL CATALYTIC V1 COMPLEX (MAIN COMPONENTS: SUBUNTS A, B,
CC C, D, E, AND F) ATTACHED TO AN INTEGRAL MEMBRANE V0 PROTON PORE
CC COMPLEX (MAIN COMPONENT: THE PROTEOLIPID PROTEIN), WHICH IS PRESENT
CC AS A HEXAMER THAT FORMS THE PROTON-CONDUCTING PORE).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. VACUOLAR.
CC -1- MISCELLANEOUS: THIS SUBUNIT BINDS DICYCLOHEXYLCARBODIIMIDE (DCDD)
CC WHICH INHIBITS THE ATPASE (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE V-ATPASE PROTEOLIPID SUBUNIT FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC
CC EMBL; AB000918; BAA22596.1; -.
CC EMBL; Z29095; CAA82355.1; -.
CC PIR; S40714; S40714.
CC HSSP; P00844; 1A91.
CC WORMPEP; R10E11.2; CE06290.
CC INTERPRO; IPR000245; -.
CC INTERPRO; IPR002379; -.
CC PFM; PF00137; ATP-synt.C; 2.
CC PRINTS; PR00122; VACATPASE.
CC KW Hydrolyase; Hydrogen ion transport; ATP synthesis; Transmembrane;
CC Multigene family.
CC FT DOMAIN 1 15 LUMENAL (POTENTIAL).
CC FT TRANSSEM 16 36 POTENTIAL.
CC FT DOMAIN 37 58 CYTOPLASMIC (POTENTIAL).
CC FT TRANSSEM 59 79 POTENTIAL.
CC FT DOMAIN 80 98 LUMENAL (POTENTIAL).
CC FT TRANSSEM 99 119 POTENTIAL.
CC FT DOMAIN 120 137 CYTOPLASMIC (POTENTIAL).
CC FT TRANSSEM 138 158 POTENTIAL.
CC FT DOMAIN 159 161 LUMENAL (POTENTIAL).
CC FT BINDING 145 145 DICYCLOHEXYLCARBODIIMIDE (POTENTIAL).
CC SO SEQUENCE 161 AA; 16409 MW; 67EDBD214F12F6E CAC64;
SQ
Query Match 89.7%; Score 26; DB 1; Length 161;
Best Local Similarity 83.3%; Pred. No. 24;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 1 MVVKGK 6
Db 77 MVVKGK 82
RESULT 6
ID MOV_P_TMYTO STANDARD; PRT; 264 AA.
AC MOV_P_TMYTO
AC P03584;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 01-APR-1993 (Rel. 25, Last annotation update)
DE MOVEMENT PROTEIN (CELL-TO-CELL TRANSPORT PROTEIN) (30 KDA PROTEIN).
OS Tobacco mosaic virus (strain tomato/L) (TMV).

```

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN 11
RP SEQUENCE FROM N.A.
RX MEDLINE: 83220776.
RA Takamatsu N., Ohno T., Meshi T., Okada Y.:
RT "Molecular cloning and nucleotide sequence of the 30K and the coat
RL protein cistron of TMV (tomato strain) genome."; Nucleic
RN Acids Res. 11:3767-3778(1983).
RP SEQUENCE FROM N.A.
RX MEDLINE: 85157522.
RA Ohno T., Aoyagi M., Yamanashi Y., Saito H., Ikawa S., Meshi T.,
Okada Y.:
RT "Nucleotide sequence of the tobacco mosaic virus (tomato strain)
RL genome and comparison with the common strain genome."; J.
CC Biochem. 96:1915-1923(1984).
CC - FUNCTION: INVOLVED IN TRANSPORT OF THE VIRUS FROM THE INITIALLY
CC INFECTED CELLS TO ADJACENT CELLS, POSSIBLY BY MODIFYING THE
CC FUNCTION OF THE PLASMODESMATA. ALSO INFLUENCES LOCAL LESION
CC DEVELOPMENT. BINDS TO SINGLE-STRANDED NUCLEIC ACID.
CC -
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X02144; CA26083.1; -
DR PIR: A04182; WMBVT3.
DR INTERPRO: IPR01022; -
DR PIR: P01107; Tobamo_MP.1.
DR PRINTS: PR00964; MOVEMENT.
KW DNA-binding; Transport.
SQ SEQUENCE 264 AA; 29291 MW; 3B01EB2359AF9C4E CRC64;

Query Match 89.7%; Score 26; DB 1; Length 264;
Best Local Similarity 83.3%; Pred. No. 39;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVKGK 6 :|||||
DB 3 LVYK GK 8

RESULT 7
MOV_P_TOMVLA STANDARD; PRT; 264 AA.
ID MOV_P_TOMVLA
AC P29809;
DT 01-APR-1993 (rel. 25, Created)
DT 01-APR-1993 (rel. 25, Last sequence update)
DT 01-APR-1993 (rel. 25, Last annotation update)
DE MOVEMENT PROTEIN (CELL-TO-CELL TRANSPORT PROTEIN) (30 KDA PROTEIN).
OS Tobacco mosaic virus (strain Lili).
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92113565.
RA Calder V.L., Palukaitis P.:
RT "Nucleotide sequence analysis of the movement genes of resistance
RL breaking strains of tomato mosaic virus."; J. Gen. Virol. 73:165-168(1992).
CC - FUNCTION: INVOLVED IN TRANSPORT OF THE VIRUS FROM THE INITIALLY
CC INFECTED CELLS TO ADJACENT CELLS, POSSIBLY BY MODIFYING THE
CC FUNCTION OF THE PLASMODESMATA. ALSO INFLUENCES LOCAL LESION
CC DEVELOPMENT. BINDS TO SINGLE-STRANDED NUCLEIC ACID.
CC PIR: J01456; WMBVL1.
DR INTERPRO: IPR01022; -
DR PIR: P01107; Tobamo_MP.1.
DR PRINTS: PR00964; MOVEMENT.
KW DNA-binding; Transport.
CC -----

SQ SEQUENCE 264 AA; 29354 MW; 16E9675B6680F391 CRC64;

Query Match 89.7%; Score 26; DB 1; Length 264;
Best Local Similarity 83.3%; Pred. No. 39;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVKGK 6 :|||||
DB 3 LVYK GK 8

RESULT 8
MOV_P_TOMVLA STANDARD; PRT; 264 AA.
ID MOV_P_TOMVLA
AC P29809;
DT 01-APR-1993 (rel. 25, Created)
DT 01-APR-1993 (rel. 25, Last sequence update)
DT 01-APR-1993 (rel. 25, Last annotation update)
DE MOVEMENT PROTEIN (CELL-TO-CELL TRANSPORT PROTEIN) (30 KDA PROTEIN).
OS Tobacco mosaic virus (strain Lili).
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 92113565.
RA Calder V.L., Palukaitis P.:
RT "Nucleotide sequence analysis of the movement genes of resistance
RL breaking strains of tomato mosaic virus."; J. Gen. Virol. 73:165-168(1992).
CC - FUNCTION: INVOLVED IN TRANSPORT OF THE VIRUS FROM THE INITIALLY
CC INFECTED CELLS TO ADJACENT CELLS, POSSIBLY BY MODIFYING THE
CC FUNCTION OF THE PLASMODESMATA. ALSO INFLUENCES LOCAL LESION
CC DEVELOPMENT. BINDS TO SINGLE-STRANDED NUCLEIC ACID.
CC PIR: J01457; WMBVL2.
DR INTERPRO: IPR01022; -
DR PIR: P01107; Tobamo_MP.1.
DR PRINTS: PR00964; MOVEMENT.
KW DNA-binding; Transport.
SQ SEQUENCE 264 AA; 29396 MW; 4737590A4EB8903B CRC64;

Query Match 89.7%; Score 26; DB 1; Length 264;
Best Local Similarity 83.3%; Pred. No. 39;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVKGK 6 :|||||
DB 3 LVYK GK 8

RESULT 9
MOV_P_TOMVLA STANDARD; PRT; 268 AA.
ID MOV_P_TOMVLA
AC P03583;
DT 21-JUL-1986 (rel. 01, Created)
DT 21-JUL-1986 (rel. 01, Last sequence update)
DT 15-DEC-1998 (rel. 37, Last annotation update)
DE MOVEMENT PROTEIN (CELL-TO-CELL TRANSPORT PROTEIN) (30 KDA PROTEIN).
OS Tobacco mosaic virus (vulgare) (TMV).
OS Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 83299880.
RA Goelet P., Lomonosoff G.P., Butler P.J.G., Akam M.E., Gait M.J.,
Karn J.:
RT "Nucleotide sequence of tobacco mosaic virus RNA."; Proc. Natl. Acad. Sci. U.S.A. 79:5818-5822(1982).
CC - FUNCTION: INVOLVED IN TRANSPORT OF THE VIRUS FROM THE INITIALLY
CC INFECTED CELLS TO ADJACENT CELLS, POSSIBLY BY MODIFYING THE
CC FUNCTION OF THE PLASMODESMATA. ALSO INFLUENCES LOCAL LESION
CC DEVELOPMENT. BINDS TO SINGLE-STRANDED NUCLEIC ACID.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration

RA Elz B., Adrian T., Ping-Akerblom P.;
 RT "Immunological adenovirus variant strains of subgenus D: comparison
 of the hekon and fiber sequences.";
 RL Virology 213:313-320(1995).
 RN [2].
 RP SEQUENCE FROM N.A.
 RC SEPARATE-ISOLATE HXB2956, AND HXH10009;
 RX MEDLINE; 95133193.
 RA Ping-Akerblom P., Adrian T.;
 RT "Characterization of adenovirus subgenus D fiber genes.";
 RL Virology 206:564-571(1995).
 CC "FUNCTION: RECOGNIZES THE CELL RECEPTOR. SERVES AS THE LIGAND
 BETWEEN THE ADENOVIRUS CAPSID AND THE HOST CELL RECEPTOR."
 CC "SUBUNIT: HOMOTRIMER."
 CC "THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch)."

CC
 CC EMBL; X72935; CAAS1440.1; -;
 CC EMBL; X72936; CAAS1441.1; -;
 CC EMBL; X74669; CAAS2733.1; -;
 CC PIR; S32663; S32663.
 CC PIR; S37278; S37278.
 CC HSP; P11818; 1K8B.
 CC INTERPRO; IPR000931; -;
 CC INTERPRO; IPR000939; -;
 CC INTERPRO; IPR000978; -;
 CC PIR; PF00608; adeno-fiber2; 2.
 CC PIR; PF00541; adeno-fiber1; 1.
 CC PRINTS; PR00307; ADENOVSFIBRE.
 CC FIBER PROTEIN.
 KW SEQUENCE 367 AA; 39983 MW; E7A4B930929C5E0C CRC64;
 SQ

Query Match 89.7%; Score 26; DB 1; Length 367;
 Best Local Similarity 83.3%; Pred. No. 52;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGVKX 6
 :|||||
 Db 220 LVKXK 225

RESULT 13
 SBPL_MOUSE
 ID SBPL_MOUSE STANDARD; PRT; 472 AA.
 AC 013228;
 DT 15-JUL-1999 (Rel. 38, Created)
 DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE SELENIUM-BINDING PROTEIN 1.
 GN SELENBP1 OR SBP.
 OS Homo sapiens (Human).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 RN [1].
 RP SEQUENCE FROM N.A.
 RC TISSUE=HEART;
 RX MEDLINE; 97179296.
 RA Chang P.W.G., Tsui S.K.W., Liew C., Lee C., Waye M.M.Y., Fung K.;
 RT "Isolation, characterization, and chromosomal mapping of a novel CDNA
 clone encoding human selenium binding protein.";
 RL J. Cell. Biochem. 64:217-224(1997).
 CC "FUNCTION: NOT KNOWN. BIND SELENIUM."
 CC "SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY)."
 CC "SIMILARITY: BELONGS TO THE SELENIUM-BINDING PROTEIN FAMILY."
 CC "THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)."

CC
 CC EMBL; U29091; AAB02395.1; -;
 CC DR MIM; 604188; -;
 KW Selenium.
 SO SEQUENCE 472 AA; 52313 MW; F484CF1CD68FC3B CRC64;

Query Match 89.7%; Score 26; DB 1; Length 472;
 Best Local Similarity 83.3%; Pred. No. 66;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGVKX 6
 :|||||
 Db 376 LVKXK 381

RESULT 14
 SBPL_MOUSE
 ID SBPL_MOUSE STANDARD; PRT; 472 AA.
 AC P17563;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE SELENIUM-BINDING PROTEIN 1 (56 KDA SELENIUM-BINDING PROTEIN) (SP56).
 GN SELENBP1 OR LPSB.
 OS Mus musculus (Mouse).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 RN [1].
 RP SEQUENCE FROM N.A.
 RC TISSUE=LIVER;
 RX MEDLINE; 91029855.
 RA Bansel M.P., Mukhopadhyay T., Scott J., Cook R.G., Mukhopadhyay R.,
 RA Medina D.;
 RT "DNA sequencing of a mouse liver protein that binds selenium:
 implications for selenium's mechanism of action in cancer
 prevention.";
 RL Carcinogenesis 11:2071-2073(1990).
 CC "FUNCTION: NOT KNOWN. BIND SELENIUM."
 CC "SUBCELLULAR LOCATION: CYTOPLASMIC."
 CC "TISSUE SPECIFICITY: HIGHLY EXPRESSED IN LIVER, KIDNEY AND, TO A
 CC LESSER EXTENT, LUNG."
 CC "SIMILARITY: BELONGS TO THE SELENIUM-BINDING PROTEIN FAMILY."
 CC "THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch)."

CC
 CC EMBL; M32032; AAA40104.1; -;
 CC DR PIR; S27878; S27878.
 CC DR SWISS-2DPAGE; P17563; MOUSE.
 CC MGD; MGI:96825; SELENBP1.
 KW Selenium.
 SO SEQUENCE 472 AA; 52352 MW; D501292C4876033D CRC64;

Query Match 89.7%; Score 26; DB 1; Length 472;
 Best Local Similarity 83.3%; Pred. No. 66;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGVKX 6
 :|||||
 Db 376 LVKXK 381

```

RESULT 15
SBP2_MOUSE
ID SBP2_MOUSE STANDARD: PRT: 472 AA.
AC 063836:
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE SELENIUM-BINDING PROTEIN 2 (56 KDA ACETAMINOPHEN-BINDING PROTEIN)
DE (AF956).
GN SELENBP2 OR LPSB2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BALB/C; TISSUE-LIVER:
RX MEDLINE: 93201669.
RA Lanfear J., Fleming J., Walker M., Harrison P.;
RT "Different patterns of regulation of the genes encoding the closely
RT related 56 kDa selenium and acetaminophen-binding proteins in normal
RT tissues and during carcinogenesis.";
RT Carcinogenesis 14:335-340(1993).
RN [2]
RP SEQUENCE OF 175-189; 196-220; 228-242; 290-296; 334-343 AND 399-408.
RX MEDLINE: 92171951.
RA Punford N.R., Martin B.M., Hinson J.A.;
RT "A metabolite of acetaminophen covalently binds to the 56 kDa
RT selenium binding protein.";
RL Biochem. Biophys. Res. Commun. 182:1348-1355(1992).
CC -1- FUNCTION: NOT KNOWN: BIND SELENIUM AND ACETAMINOPHEN.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- TISSUE SPECIFICITY: MAINLY EXPRESSED IN LIVER.
CC -1- SIMILARITY: BELONGS TO THE SELENIUM-BINDING PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: S56599; AAB25841.2; -
DR MGD: MGI:104859; SELENBP2.
KW Selenium.
SQ SEQUENCE 472 AA; 52628 MW; C32FE819C4AD07CA CRC64;

```

```

Query Match          89.78; Score 26; DB 1; Length 472;
Best Local Similarity 83.38; Pred. No. 66;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
OY 1 MYVKGK 6
   :|||||
DB 376 LVVKGK 381

```

Search completed: December 14, 2000, 10:14:59
Job time: 224 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 14, 2000, 10:12:27 ; Search time 40.16 Seconds

(Without alignments)
13.950 Million cell updates/sec

Title: US-09-051-034a-13

Sequence: 1 MYVKGK 6

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 297973 seqs, 93374136 residues

Total number of hits satisfying chosen parameters: 297973

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

1: SP. Archaea: *
2: SP. Bacteria: *
3: SP. Fungi: *
4: SP. Human: *
5: SP. Invertebrate: *
6: SP. Mammal: *
7: SP. MHC: *
8: SP. Organelle: *
9: SP. Phage: *
10: SP. Plant: *
11: SP. Rodent: *
12: SP. Virus: *
13: SP. Vertebrate: *
14: SP. Unclassified: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	29	100.0	327	10	Q9S260 arabidopsi
2	28	96.6	162	2	Q67338 aquifex aeo
3	27	93.1	633	1	Q59176 pyrococcus
4	27	93.1	662	5	Q62139 arabidopsi
5	26	89.7	242	12	Q9WNG4 tobacc
6	26	89.7	264	12	Q83481 tobacc
7	26	89.7	264	12	Q9YJ09 tobacc
8	26	89.7	268	12	Q98746 tobacc
9	26	89.7	268	12	Q91775 tobacc
10	26	89.7	268	12	Q9QPN6 tobacc
11	26	89.7	333	10	Q82739 arabidopsi
12	26	89.7	343	11	Q88537 mus musculi
13	26	89.7	502	2	Q9RGV6 salmone
14	26	89.7	815	2	Q32178 bacilli
15	25	86.2	138	1	Q9URF5 pyrococcus
16	25	86.2	163	1	Q28651 archaeglob
17	25	86.2	175	3	Q94390 schistosach
18	25	86.2	179	5	Q22577 caenorhabdi
19	25	86.2	191	5	Q9XTW9 meloidogyn

20	25	86.2	204	1	Q29462 archaeglob
21	25	86.2	231	1	Q58733 methanococ
22	25	86.2	336	5	Q9Y073 lymnaea sta
23	25	86.2	352	1	Q9UZG0 pyrococcus
24	25	86.2	367	2	Q9XB11 bacillus ce
25	25	86.2	372	1	Q9V2A2 pyrococcus
26	25	86.2	387	2	Q9Z111 streptococ
27	25	86.2	395	2	Q67558 aquifex aeo
28	25	86.2	420	1	Q9YFE1 aeropyrum p
29	25	86.2	444	3	Q9Y838 mycosphere
30	25	86.2	446	1	Q9Y2V8 thermoprote
31	25	86.2	455	10	P93094 cucumis mel
32	25	86.2	458	8	Q03078 ophioglossu
33	25	86.2	472	10	Q9ZRC0 arabidopsi
34	25	86.2	548	5	Q18041 caenorhabdi
35	25	86.2	583	5	Q9XZ66 caenorhabdi
36	25	86.2	589	2	Q9S158 commomonas t
37	25	86.2	700	3	Q74935 yarrowia li
38	25	86.2	774	5	Q9XZ53 drosophila
39	25	86.2	808	12	Q9QR00 canine coro
40	25	86.2	814	5	Q9VEJ2 drosophila
41	25	86.2	884	5	Q17429 caenorhabdi
42	25	86.2	910	4	Q9Y620 homo sapien
43	25	86.2	912	5	Q76283 trypanosoma
44	25	86.2	1098	12	P89285 xestia c-ni
45	25	86.2	1262	5	Q20684 caenorhabdi

ALIGNMENTS

RESULT 1

ID	Q9S260	PRELIMINARY:	PRT:	327 AA.
AC	Q9S260:			
DT	01-MAY-2000 (TrEMBLrel. 13, Created)			
DT	01-MAY-2000 (TrEMBLrel. 13, Last sequence update)			
DT	01-MAY-2000 (TrEMBLrel. 13, Last annotation update)			
DE	HYPOHETICAL 35.8 KDA PROTEIN.			
GN	FL6J13.20.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;			
OC	Magnoliophyta; eudicotyledons; Rosidae; eurosids II; Brassicales;			
OC	Brassicaceae; Arabidopsis.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Bevan M., Hilbert H., Braun M., Holzer E., Brandt A., Duesterhoeft A.,			
RA	Bancroft I., Mewes H.W., Mayer K.F.X., Schueller C.;			
RL	Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.			
RL	Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	EU Arabidopsis sequencing project;			
RA	Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AL049638; CAB40936.1; -			
KW	HYPOTHETICAL protein.			
SO	SEQUENCE 327 AA; 35845 MW; B2BB47DF4A83E635 CRC64;			

Query Match 100.0%; Score 29; DB 10; Length 327;
Best Local Similarity 100.0%; Pred. No. 64;
Matches 6; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYVKGK 6

DB 275 MYVKGK 280

RESULT 2

ID	Q67338	PRELIMINARY:	PRT:	162 AA.
AC	Q67338:			
DT	01-AUG-1998 (TrEMBLrel. 07, Created)			
DT	01-AUG-1998 (TrEMBLrel. 07, Last sequence update)			

DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
 DE NADH DEHYDROGENASE I CHAIN J.
 GN NUDJ1.
 OS Aquifex aeolicus.
 OC Bacteria; Aquificales; Aquificaceae; Aquifex.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VF5;
 RA MEDLINE: 98196666.
 RA Decker G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
 RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Anjay M., Huber R.,
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.,
 RT The complete genome of the hyperthermophilic bacterium Aquifex
 RT aeolicus.
 RL Nature 392:353-358(1998).
 DR EMBL: AE000734; AAC07301.1; -
 DR INTERPRO: IPR001457; -
 DR PFAM: PF00499; oxidored_3; 1.
 SQ SEQUENCE 162 AA; 17859 MW; F45EDPB206E3AF8E CRC64;

Query Match 96.6%; Score 28; DB 2; Length 162;
 Best Local Similarity 83.3%; Pred. No. 52;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MVVKG 6
 DB 108 MVKKG 113

RESULT 3
 ID 059176 PRELIMINARY; PRT; 633 AA.
 AC 059176;
 DT 01-AUG-1998 (TReMBLrel. 07, Created)
 DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
 DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)
 DE HYPOTHELTICAL 73.4 KDA PROTEIN PH1507.
 GN PH1507.
 OS Pyrococcus horikoshii.
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-OT3;
 RA MEDLINE: 98344137.
 RA Kavarabavasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
 RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
 RA Sawai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohnuki Y.,
 RA Buhanaishi T., Tanaka T., Kudo Y., Yamazaki J., Kushiida N., Oguchi A.,
 RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
 RA Masuchi Y., Shizuya H., Kikuchi H.;
 RT Complete sequence and gene organization of the genome of a hyper-
 RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.
 RL Nature 392:555-561(1998).
 DR EMBL: AP000006; BAA50615.1; -
 DR INTERPRO: IPR001617; -
 DR PFAM: PF00005; ABC_tran; 1.
 KW Hypothetical protein.
 SQ SEQUENCE 633 AA; 73382 MW; 89ABF4FE807B5C42 CRC64;

Query Match 93.1%; Score 27; DB 1; Length 633;
 Best Local Similarity 66.7%; Pred. No. 3.7e+02;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 MVVKG 6
 DB 1 MVKKG 6

RESULT 4
 ID 062139 PRELIMINARY; PRT; 662 AA.

AC 062139;
 DT 01-AUG-1998 (TReMBLrel. 07, Created)
 DT 01-AUG-1998 (TReMBLrel. 07, Last sequence update)
 DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
 DE F08A8.4 PROTEIN.
 GN F08A8.4.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Harris B.;
 RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE: 94150718.
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
 RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
 RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
 RA Lightning J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
 RA Parsons J., Percy C., Riffen L., Roopra A., Saunders D., Showkneen R.,
 RA Smalton N., Smith A., Sonhammer E., Staden R., Sulston J.,
 RA Thierly-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
 RA Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 RT elegans".
 RL Nature 368:32-38(1994).
 DR EMBL: Z99710; CAB16866.1; -
 DR INTERPRO: IPR02655; -
 DR PFAM: PF01756; ACOX; 1.
 SQ SEQUENCE 662 AA; 74770 MW; B1EA46FAE920BD68 CRC64;

Query Match 93.1%; Score 27; DB 5; Length 662;
 Best Local Similarity 66.7%; Pred. No. 3.9e+02;
 Matches 4; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 MVVKG 6
 DB 205 MVKKG 210

RESULT 5
 ID 09WNG4 PRELIMINARY; PRT; 242 AA.
 AC 09WNG4;
 DT 01-NOV-1999 (TReMBLrel. 12, Created)
 DT 01-NOV-1999 (TReMBLrel. 12, Last sequence update)
 DT 01-MAY-2000 (TReMBLrel. 13, Last annotation update)
 DE 26.8 KDA PROTEIN.
 GN MP.
 OS Tobacco mosaic virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Sheng O.B., Gong Y.;
 RT "Cloning and sequence analysis of infectious cDNA of attenuated
 RT tobacco mosaic virus".
 RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF155507; AAD44328.1; -
 DR INTERPRO: IPR01022; -
 DR PFAM: PF01107; Tobamo_MP; 1.
 DR PRINTS: PRO0964; MOVEMENT.
 SQ SEQUENCE 242 AA; 26802 MW; E8A8C91183A3FD03 CRC64;

Query Match 89.7%; Score 26; DB 12; Length 242;
 Best Local Similarity 83.3%; Pred. No. 2.3e+02;
 Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MVVKG 6

Db :|||||
3 LVYK GK 8

RESULT 6
ID 083481 PRELIMINARY; PRT; 264 AA.
AC 083481;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TEMBLrel. 13, Last annotation update)
DE 30K PROTEIN.
OS Tobacco mosaic virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-2A;
RA Ikeda R., Watanabe Y., Okada Y.;
RT "Mutation in the tobacco mosaic virus 30-KD protein gene overcome Tm-2a resistance in tomato."
RL Submitted (AUG-1993) to the EMBL/GenBank/DBJ databases.
DR EMBL: D17458; BAA04268.1; -.
DR INTERPRO: IPR001022; -.
DR PFAM: PF01107; Tobamo_MP. 1.
DR PRINTS: PRO0964; MOVEMENT.
SQ SEQUENCE 264 AA; 29362 MW; 36B7EA334F1939D8 CRC64;

Query Match 89.7%; Score 26; DB 12; Length 264;
Best Local Similarity 83.3%; Pred. No. 2.5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVYK GK 6
:|||||
Db 3 LVYK GK 8

RESULT 7
ID 09YJ09 PRELIMINARY; PRT; 264 AA.
AC 09YJ09;
DT 01-MAY-1999 (TEMBLrel. 10, Created)
DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
DT 01-MAY-2000 (TEMBLrel. 13, Last annotation update)
DE MOVEMENT PROTEIN.
OS Tobacco mosaic virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-S-1;
RA Zhou X., Xue C., Chen Q., Qi Y., Li D.;
RT "Complete nucleotide sequence of a Chinese isolate of tomato mosaic virus."
RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE OF 1-200 FROM N.A.
RC STRAIN-S1;
RA Zhou X., Xue C., Qi Y., Li D.;
RT "Isolation and nucleotide sequence of the 30k and the coat protein cistron of a Chinese isolate of tomato mosaic virus."
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: A1132845; CAB36999.1; -.
DR EMBL: A011934; CAA09878.1; -.
DR INTERPRO: IPR001022; -.
DR PFAM: PF01107; Tobamo_MP. 1.
DR PRINTS: PRO0964; MOVEMENT.
SQ SEQUENCE 264 AA; 29261 MW; 018394057CA46FAB CRC64;

Query Match 89.7%; Score 26; DB 12; Length 264;
Best Local Similarity 83.3%; Pred. No. 2.5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVYK GK 6
:|||||
Db 3 LVYK GK 8

RESULT 8
ID 098746 PRELIMINARY; PRT; 268 AA.
AC 098746;
DT 01-FEB-1997 (TEMBLrel. 02, Created)
DT 01-FEB-1997 (TEMBLrel. 02, Last sequence update)
DT 01-MAY-2000 (TEMBLrel. 13, Last annotation update)
DE 30K PROTEIN.
OS Tobacco mosaic virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RAKKYO;
RX MEDLINE: 96265021.
RA Chen J., Watanabe Y., Sako N., Ohshima K., Okada Y.;
RT "Complete nucleotide sequence and synthesis of infectious in vitro transcripts from a full-length cDNA clone of a Rakkyo strain of tobacco mosaic virus."
RL Arch. Virol. 141:885-900(1996).
DR EMBL: D63809; BAA09878.1; -.
DR INTERPRO: IPR001022; -.
DR PFAM: PF01107; Tobamo_MP. 1.
DR PRINTS: PRO0964; MOVEMENT.
SQ SEQUENCE 268 AA; 29849 MW; 2E843BE774DF312F CRC64;

Query Match 89.7%; Score 26; DB 12; Length 268;
Best Local Similarity 83.3%; Pred. No. 2.5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MVYK GK 6
:|||||
Db 3 LVYK GK 8

RESULT 9
ID 091275 PRELIMINARY; PRT; 268 AA.
AC 091275;
DT 01-NOV-1998 (TEMBLrel. 08, Created)
DT 01-NOV-1998 (TEMBLrel. 08, Last sequence update)
DT 01-MAY-2000 (TEMBLrel. 13, Last annotation update)
DE MOVEMENT PROTEIN.
OS Tobacco mosaic virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-B;
RA Zhou X., Chen Q., Xue Z., Li D.;
RT "Sequence of movement protein gene of tobacco mosaic virus faba bean strain."
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-B935A;
RA Xue C., Zhou X., Chen Q., Qi Y., Li D.;
RT "Complete nucleotide sequence and genome organization of tobacco mosaic virus isolated from Vicia faba."
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: A006891; CA07412.1; -.
DR EMBL: A011933; CAA09876.1; -.
DR INTERPRO: IPR001022; -.
DR PFAM: PF01107; Tobamo_MP. 1.
DR PRINTS: PRO0964; MOVEMENT.
SQ SEQUENCE 268 AA; 30012 MW; 8E6087C7789A4464 CRC64;

Query Match 89.7%; Score 26; DB 12; Length 268;

Best Local Similarity 83.3%; Pred. No. 2.5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MVMGK 6
:|||||
Db 3 LVVGK 8

RESULT 10

090EN6 ID 090PN6 PRELIMINARY; PRT: 268 AA.
AC 090PN6
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE MOVEMENT PROTEIN.
GN P.
OS Tobacco mosaic virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
RN [1]
RP SEQUENCE FROM N.A.
RA Sheng Q.B., Gong Y.;
RT Cloning and sequence analysis of infectious cDNA."
RL Submitted (JUL-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF165190; AAD47819.1; -.
DR INTERPRO: IPR001022; -.
DR PFAM: PF01107; Tobacco_MP; 1.
DR PRINTS: PR00964; MOVEMENT.
SQ SEQUENCE 268 AA; 29942 MW; A50D311C45399E59 CRC64;

Query Match 89.7%; Score 26; DB 12; Length 268;
Best Local Similarity 83.3%; Pred. No. 2.5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MVMGK 6
:|||||
Db 3 LVVGK 8

RESULT 11

082739 ID 082739 PRELIMINARY; PRT: 333 AA.
AC 082739
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE THEORETICAL 36.4 KDA PROTEIN.
GN P19.80.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
OC Magnoliophyta; eudicotyledons; Rosidae; eucosids II; Brassicales;
OC Brassicaceae; Arabidopsis.
RN [1]
RP SEQUENCE FROM N.A.
RA Peters S.A., van Staveren M., Dirkse W., Stiekema W., Bancroft I.,
RT Characterization of the trn2 region of the Inchi1 plasmid R27."
RL Submitted (JUL-1998) to the EMBL/Genbank/DBJ databases.
DR EMBL: AL031018; CAA19805.1; -.
KW Hypothetical protein.
SQ SEQUENCE 333 AA; 36350 MW; A70BA7FB93F9EDB CRC64;

Query Match 89.7%; Score 26; DB 10; Length 333;
Best Local Similarity 83.3%; Pred. No. 3.2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MVMGK 6
:|||||
Db 272 LVVGK 277

RESULT 12

088537 ID 088537 PRELIMINARY; PRT: 343 AA.
AC 088537
DT 01-NOV-1998 (TREMblrel. 08, Created)
DT 01-NOV-1998 (TREMblrel. 08, Last sequence update)
DT 01-JUN-2000 (TREMblrel. 14, Last annotation update)
DE N-FORMYLPEPTIDE RECEPTOR-LIKE 3.
GN FPR-RS3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
RN [1]
RP SEQUENCE FROM N.A.
RA Gao J.-L., Chen H., Filie J.D., Kozak C.A., Murphy P.M.;
RT "Differential expansion of the N-formylpeptide receptor gene cluster
in human and mouse."
RL Genomics 51:270-276(1998).
DR EMBL: AF071181; AAC34586.1; -.
DR INTERPRO: IPR000276; -.
DR PFAM: PF00001; 7tm.1; 1.
DR PRINTS: PR00237; GPCRHHODPSN.
DR PRINTS: PR00526; FMETLEUPHER.
DR PROSITE: PS00237; G_PROTEIN_RECEPTOR; UNKNOWN_1.
SQ SEQUENCE 343 AA; 38111 MW; 0F726BF195CBFB29 CRC64;

Query Match 89.7%; Score 26; DB 11; Length 343;
Best Local Similarity 83.3%; Pred. No. 3.3e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MVMGK 6
:|||||
Db 85 MVMGK 90

RESULT 13

09RGV6 ID 09RGV6 PRELIMINARY; PRT: 502 AA.
AC 09RGV6
DT 01-MAY-2000 (TREMblrel. 13, Created)
DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMblrel. 13, Last annotation update)
DE TRHW.
GN TRHW.
OS Salmonella typhi.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 99296679.
RA Rooker M.W., Sherburne C., Lawley T.D., Taylor D.E.;
RT Characterization of the trn2 region of the Inchi1 plasmid R27."
RL Plasmid 41:226-239(1999).
DR EMBL: AF105019; AAD54026.1; -.
KW Plasmid.
SQ SEQUENCE 502 AA; 56512 MW; F2FFE73D12B98FB6 CRC64;

Query Match 89.7%; Score 26; DB 2; Length 502;
Best Local Similarity 83.3%; Pred. No. 5e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MVMGK 6
:|||||
Db 242 LVVGK 247

RESULT 14

032178 ID 032178 PRELIMINARY; PRT: 815 AA.

AC 032178;
DT 01-JAN-1998 (TReMBLrel. 05, Created)
DT 01-JAN-1998 (TReMBLrel. 05, Last sequence update)
DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)
DE YUSL PROTEIN.
GN YUSL.
OS Bacillus subtilis.
OC Bacteria: Firmicutes: Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-168;
RX MEDLINE: 98044033.
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borcher S.,
RA Borriss R., Boutsier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Enlian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Guim S.Y., Glaser P., Goffeau A., Gollighly E.J., Grandi G.,
RA Guseppl G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porvollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Ray M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Serior P., Shin B.S., Soldo B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Toesio V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambut R., Wedler E., Wedler H., Weltzenegger T.,
RA Winters P., Wiput A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the gram-positive bacterium Bacillus
RT subtilis".
RT Nature 390:249-256(1997).
RL [2]
RN SEQUENCE FROM N.A.
RP STRAIN-168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: 299120; CAB5273.1; -.
DR INTERPRO: IPR000205; -.
DR INTERPRO: IPR001753; -.
DR INTERPRO: IPR002135; -.
DR PFM: PFM0378; ECH; 1.
DR PFM: PFM0725; 3HCDH; 1.
SQ SEQUENCE 815 AA: 90148 MW: 5415086EF48748C1 CRC64;

Query Match 89.7%; Score 26; DB 2; Length 815;
Best Local Similarity 83.3%; Pred. No. 8.3e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MYVKGK 6
DB 806 MLVKGK 811

RESULT 15
O9UYR5 PRELIMINARY; PRT; 138 AA.
AC O9UYR5;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)

DT 01-JUN-2000 (TReMBLrel. 14, Last annotation update)
DE HYPOTHETICAL 15.9 KDA PROTEIN.
GN PAB0960.
OS Pyrococcus abyssi.
OC Archaea: Euryarchaeota: Thermococcales; Thermococcaceae; Pyrococcus.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-ORSAV;
RA Heilig R.;
RT "Pyrococcus abyssi genome sequence: Insights into archaeal chromosome
RT structure and evolution."
RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ248287; CAB50347.1; -.
DR INTERPRO: IPR002145; -.
DR PFM: PFM1402; HTH 4; 1.
KW Hypothetical protein.
SQ SEQUENCE 138 AA: 15941 MW: 008B785364DBB302 CRC64;

Query Match 86.2%; Score 25; DB 1; Length 138;
Best Local Similarity 83.3%; Pred. No. 2.2e+02;
Matches 5; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MYVKGK 6
DB 101 IYVKGK 106

Search completed: December 14, 2000, 10:12:30
Job time: 124 sec

